Ein aufmerksamkeitsgestütztes, biologienahes Objekt-Erkennungs- und Verfolgungssystem mit impulscodierenden Neuronen

Dissertation
zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich Physik
der Philipps-Universität Marburg
vorgelegt von

Martin Lothar Pauly

aus Karlsruhe

Marburg/Lahn, März 2000

Vom Fachbereich Physik der Philipps-Universität als Dissertation angenommen
am 11. Mai 2000
Erstgutachter: Prof. Dr. R. Eckhorn
Zweitgutachter: Prof. Dr. F. Rösler
Tag der mündlichen Prüfung: 2. Juni 2000

Inhaltsverzeichnis

Zusammenfassung

Ziel. Die vorliegende Arbeit präsentiert ein neuronales Netz zur Blicksteuerung, das das Auffinden sowie die Verfolgung von Objekten in bewegten realen Szenen ermöglicht. Im Gegensatz zu den meisten technischen Lösungen zur Objektverfolgung ist explizites Objektwissen nicht erforderlich; die Identifikation von Objekten geschieht ausschließlich aufgrund von Merkmalskontrasten zum Hintergrund (Grauwert- und Bewegungskontrast).

Aufbau des Systems. Das gesamte System ist aus retinotop angeordneten Schichten von impulscodierenden *Marburger Modellneuronen* aufgebaut und lehnt sich in seinem Aufbau stark an die bekannten Gegebenheiten im *Superior Colliculus* an; dieser steuert im Gehirn von Säugetieren die *sakkadischen Augenbewegungen*. Die ebenfalls aus der Biologie bekannten *glatten Folgebewegungen* wurden unter Verwendung desselben Netzwerks zusätzlich implementiert.

Das eigentliche System besteht aus drei Teilen, die als geschlossene Schleife betrieben werden: Vorverarbeitung, Aufmerksamkeitsschicht und Blicksteuerung. Die Vorverarbeitung extrahiert aus der Eingangssequenz (monokulare Grauwertbilder) die erwähnten Merkmalskontraste und gibt sie als retinotop angeordnete Aktivitätsverteilung an die Aufmerksamkeitsschicht weiter.

Auswahl von Blickzielen. Die Aufmerksamkeitsschicht implementiert durch eine Kombination aus lokaler Exzitation und globaler Inhibition einen ständigen *Winner-Take-All*-Wettbewerb zwischen möglichen Blickzielen: Während die unmittelbare Umgebung eines einmal angeregten Ortes durch die lokale Exzitation weiteren Input erhält, inhibieren sich alle weit voneinander entfernten Orte gegenseitig [Amari, 1977; Kopecz und Schöner, 1995]. In Verbindung mit der Feuerschwelle im *Marburger Modellneuron* führt dies dazu, daß nur kurzzeitig Aktivität an mehreren Orten der Aufmerksamkeitsschicht existieren kann. Am Ende des Wettbewerbs sind nur noch Neurone an einem Ort aktiv; dieser markiert das prominenteste Blickziel.

Weicht das so ermittelte Blickziel nur wenig von der aktuellen Blickrichtung ab, so wird der Blick kontinuierlich nachgeführt. Das Ergebnis ist beim ruhenden Objekt eine dauerhafte Fixation, beim bewegten Objekt eine *gleichmäßige*, schlupfbehaftete *Folgebewegung*. Da die Nachführung der Blickrichtung sich auf den Input des Systems auswirkt (zur Kamerabewegung entgegengesetzte Scheinbewegung der Eingangsbilder), arbeitet das gesamte System in diesem Zustand als geschlossene Regelschleife. Eine näherungsweise analytische Behandlung dieser Regelung wird angegeben.

Überschreitet die Abweichung zwischen markiertem Blickziel und aktueller Blickrichtung einen bestimmten, einstellbaren Wert, dann wird eine *Sakkade* zum Blickziel hin ausgelöst, d.h. dieses sofort fixiert. Nach der Sakkade wird der visuelle Input für 50 ms unterdrückt, um eine Relaxation der Neurone zu ermöglichen (sakkadische Suppression), anschließend baut sich die Aktivität neu auf.

Ergebnisse und Schlußfolgerungen. Bei der Anwendung auf bewegte reale Szenen (z.B. Verkehrsszenen) zeigt sich, daß das System eine sinnvolle Blicksteuerung auf der Basis der detektierten Merkmalskontraste leisten kann. Im Vergleich mit modellbasierten technischen Lösungen weist die Verfolgung bewegter Objekte aufgrund des fehlenden Modellwissens erheblich größere Ungenauigkeiten auf. Diesem Nachteil steht als Vorteil die Fähigkeit gegenüber, beliebige Objekte aufzufinden bzw. zu verfolgen, sobald diese sich visuell vom Hintergrund abheben.

In weiteren Simulationen zur *Objekt-Hintergrund-Segmentierung im Zeitbereich* (entsprechend der *Synchronisationshypothese* der Hirnforschung) wird demonstriert, daß die vorgestellte Aufmerksamkeitssteuerung diese Aufgabe in zweifacher Hinsicht erleichtern kann: 1. Durch das Auffinden und die gezielte Bearbeitung relevanter Bildausschnitte läßt sich die Komplexität des Problems deutlich reduzieren. 2. Durch *aufmerksamkeitsinduzierte Latenzen*, wie sie auch aus der Psychophysik bekannt sind, kann die zeitliche Segmentierung einer Szene erheblich erleichtert werden.

1 Einleitung

1.1 Allgemeine Einordnung

In den letzten drei Jahrzehnten ist ein Gebiet in den Mittelpunkt des Interesses der naturwissenschaftlichen Forschung gerückt, das früher aus mehreren Gründen äußerst unzugänglich erschien: Die Informationsverarbeitung in natürlichen Systemen, speziell im Nervensystem von Menschen und Tieren, erweist sich als ein faszinierendes und spannendes Forschungsgebiet, das die Grenzen der traditionellen naturwissenschaftlichen Disziplinen an fast allen entscheidenden Punkten überschreitet.

Die Anfänge einer systematischen Forschung auf diesem Gebiet reichen bis ins 19. Jahrhundert und weiter zurück. Aus dieser Zeit stammt die Begründung der Wahrnehmungspsychologie oder *Psychophysik* durch Fechner [1860], James [1890] u.a. Die *Gestaltpsychologie* von Wertheimer [1912], Köhler [1924], Metzger [1936] u.a. lieferte weitere empirische Erkenntnisse über die Eigenheiten der Wahrnehmung, die zu großen Teilen heute noch gültig sind. Aus dieser Zeit stammt auch die Vermutung, daß die Integration und Analyse sensorischer Information, also auch die Wahrnehmung, vom Zentralnervensystem (ZNS) geleistet werden.

Die moderne Neurowissenschaft hat große Anstrengungen unternommen, die komplexen Vorgänge im Nervensystem auf mikroskopischer, d.h. molekularer und biophysikalischer Ebene zu verstehen. Dies ist teilweise gelungen; so wurde z.B. in den 50er Jahren der Ablauf von Aktionspotentialen aufgeklärt [Hodgkin und Huxley, 1952], ebenso die molekularen Wirkungsmechanismen vieler chemischer Botenstoffe.

Trotzdem hat das zunehmende Wissen über die molekularen Vorgänge nur in Teilbereichen zu einem tieferen Verständnis der Informationsverarbeitung im Nervensystem beigetragen. So gibt das Wissen um die biophysikalischen Abläufe in einer Synapse für sich allein noch keinen Hinweis darauf, welche Informationen über diese Synapse übertragen werden. Auf der anderen Seite kann eine Modellierung auf der Ebene abstrakter Subsysteme ebensowenig detaillierte Einblicke in ein mögliches Zusammenspiel der Neurone bei der Informationsverarbeitung geben, selbst wenn sie eine korrekte quantitative Beschreibung der beobachteten Phänomene liefert (vgl. z.B. die in Kap. 1.4 zitierte Literatur zur Entstehung von Sakkaden).

Als vielversprechender Ansatz zur Lösung dieses Dilemmas hat sich in den letzten zwei Jahrzehnten die Methode der 'neuronalen Netze' etabliert. Die Beschreibungsebene liegt hierbei zwischen den beiden genannten Extremen; die kleinste betrachtete Einheit ist ein mehr oder weniger stark vereinfachtes Modell eines realen Neurons. Das betrachtete System besteht aus vielen solcher Modellneurone (die sich fast immer nichtlinear verhalten) und ist daher formal als nichtlineares dynamisches System aufzufassen. Inhaltlich

liegt diesem Modellierungsansatz die Hypothese zugrunde, daß sich alle wichtigen Prozesse der Informationsverarbeitung auf der Beschreibungsebene der Verschaltungsstruktur und der Systemdynamik erfassen lassen.

Für die vorliegende Arbeit ist dabei die Unterscheidung zwischen kontinuierlichen und impulscodierenden Neuronenmodellen von Bedeutung: Beim ersten Typ geht man davon aus, daß sich die Ausgangsaktivität von Neuronen ausreichend genau durch deren mittlere Pulsrate (genauer: zeitliche Impulswahrscheinlichkeitsdichte) als kontinuierliche Ausgangsgröße beschreiben läßt. Der zweite Modelltyp löst die Zeitstruktur der neuronalen Ausgangsaktivität explizit auf; diese besteht dann aus einzelnen Aktionspotentialen mit genau definiertem Auftretenszeitpunkt. Das erste derartige Neuronenmodell war das *Integrate-and-fire*-Modell von French und Stein [1970]. Das *Spike-Response*-Modell von Gerstner et al. [1991] stellen ebenso wie das in dieser Arbeit verwendete *Marburger Modellneuron* [Eckhorn et al., 1990] Erweiterungen dieses Modells dar.

Einen starken Aufschwung erfuhr dieser Modellierungsansatz mit dem Aufkommen erschwinglicher Computer und ihrer raschen Weiterentwicklung in den letzten zwei Jahrzehnten. Damit wurde es zum erstenmal möglich, komplexe Systeme, die sich der analytischen Beschreibung aufgrund von Nichtlinearitäten weitgehend entziehen (also auch neuronale Netze), in großem Umfang numerisch zu behandeln. Daraus entwickelte sich als neue wissenschaftliche Arbeitstechnik die *Simulation*: Ein komplexes (quantitativ durch Differentialgleichungen beschriebenes) System wird im Computer 'nachgebaut', so daß sich sein Verhalten unter kontrollierten Bedingungen studieren läßt.

Netzwerkmodelle neuronaler Vorgänge haben aus sich heraus keine Beweiskraft im naturwissenschaftlichen Sinn. Ihren besonderen Wert beziehen sie daraus, daß die 'Modellsprache' der neuronalen Verschaltungen im Prinzip einen direkten Vergleich mit dem biologischen System ermöglicht – auch wenn dieser oft auf große methodische Schwierigkeiten stößt. Ein zweiter wichtiger Punkt ist die Möglichkeit, am Modell die Bedingungen für eine bestimmte Funktionalität zu untersuchen und in begrenztem Umfang Vorhersagen über mögliche Verschaltungen im realen System zu treffen. Dieser Aspekt spielt in der vorliegenden Arbeit die zentrale Rolle.

1.2 Motivation und Ziel der Arbeit

Die vorliegende Arbeit entstand im Rahmen des BMBF-Projekts *Das Elektronische Auge*. Ein Ziel dieses Projekts ist es, biologisch inspirierte Modelle des menschlichen Sehvorgangs für die technische Bildverarbeitung zu nutzen. Die Motivation für diese Verbindung ist die bisher unerreichte Robustheit und Vielseitigkeit von Lebewesen bei der Verarbeitung und Umsetzung visueller Informationen. Zwar kann die technische Bildverarbeitung mittlerweile in vielen Gebieten anwendungsreife Lösungen vorweisen. Dabei handelt es sich jedoch zum größten Teil um Speziallösungen für relativ eng umgrenzte Probleme, vgl. die kurze Übersicht zur *Objekt-Hintergrund-Segmentierung* und *Objektverfolgung* in Kap. 1.4.2.

In dieser Arbeit versuche ich, das biologisch begründete Konzept der *Aufmerksamkeit* für die Objekt-Hintergrund-Segmentierung umzusetzen und nutzbar zu machen. Aufmerksamkeit im Sinn einer räumlichen Ausrichtung der Wahrnehmung spielt für Lebewesen

offenbar seit der frühen Entwicklungsgeschichte eine wichtige Rolle: Fast alle Wirbeltiere können durch die Bewegung von Augen und Kopf ihre visuelle Wahrnehmung gezielt ausrichten. Diese Ausrichtung kann sowohl durch visuelle Reize als auch durch Reize anderer Sinnesmodalitäten (akustisch, somatosensorisch) oder durch Intention ausgelöst werden. Darüber hinaus ist eine Bewegung von Augen bzw. Kopf nicht zwingend erforderlich – zumindest eine beschleunigte Verarbeitung bestimmter Teilbereiche des Sehraums kann offenbar allein durch neuronale Prozesse bewirkt werden (vgl. Kap. 2.5.2).

Von großer Bedeutung für die entwicklungsgeschichtliche Selektion dürfte eine möglichst schnelle und direkte Anbindung der motorischen Reaktionen an die Aufmerksamkeitsprozesse sein; nur so läßt sich die 'gerichtete Wahrnehmung' in adäquates Verhalten umsetzen. Die motorische Komponente wird in dieser Arbeit nur insoweit berücksichtigt, als sie die Bewegung von Auge bzw. Kamera selbst betrifft.

Bei der Modellierung der Blickbewegungen greife ich auf die funktionale Unterscheidung zwischen *Sakkaden* (plötzlichen Sprüngen der Augenposition) und kontinuierlichen *Folgebewegungen* zurück, wie sie in der Wahrnehmungsforschung seit den Arbeiten von Dodge und Cline [1901] bekannt ist. Beide Arten von Augenbewegungen behandle ich in einem einheitlichen Modell, das in seinem Aufbau stark an neurobiologischen und psychophysischen Erkenntnissen orientiert ist.

Ein Teil der hier vorgestellten Ergebnisse wurde bereits veröffentlicht, vgl. [Pauly et al., 1997, 1998, 1999; Mohraz et al., 1997].

1.3 Überblick

Im folgenden Abschnitt referiere ich kurz den derzeitigen Stand der Forschung zum Thema Augenbewegungen und Objekt-Hintergrund-Segmentierung. An eine kurze Darstellung der biologischen Grundlagen in Kap. 2 schließt sich die Beschreibung der verwendeten Modelle an: Auf der Ebene des Einzelneurons ist dies das *Marburger Modellneuron*. Auf der Netzwerkebene wird in Kap. 3 ein System vorgestellt, das die funktionale Architektur des menschlichen Sehsystems teilweise nachbildet, wobei die Unterteilung in ein langsames *Kontur-Form-System* (parvozellulärer Pfad) und ein schnelles *Transientensystem* (magnozellulärer Pfad) eine zentrale Rolle spielt.[1] Kap. 3.5 ist dem eigentlichen Aufmerksamkeits- bzw. Blicksteuerungsmechanismus gewidmet; Arbeitweise und Eigenschaften werden detailliert besprochen.

An diese Vorstellung der Teilsysteme schließen sich mit Kap. 4 Überlegungen und Grundlagensimulationen zur Segmentierung im Zeitbereich an. Kap. 5 schließlich präsentiert die Simulationsergebnisse mit realen Szenen.

Während der Simulationen mit dem Kontur-Form-System entstand als zusätzliches Ergebnis eine Grundlagenbetrachtung zur Wirkung verschiedener Typen der Nachbarschaftskopplung von Kantendetektoren, insbesondere bei verrauschten bzw. gestörten Bildern. Diese wird in Kap. 6 vorgestellt; sie ermöglicht in stationärer Näherung eine quantitative Behandlung dieser Architektur in Termen von Fehlerwahrscheinlichkeiten.

[1] Dieses System wurde zum größten Teil aus den Arbeiten von Schott [1999] und Weitzel [1998b] übernommen, die im gleichen Projekt entstanden. Alle übernommenen Teile sind im Text entsprechend gekennzeichnet, ebenso Abbildungen aus den genannten Arbeiten.

Abschließend bewerte ich das Verhalten des Gesamtsystems und seine Relevanz sowohl für die biologische Grundlagenforschung als auch für die technische Umsetzung.

1.4 Bisherige Arbeiten

1.4.1 Augenbewegungen und Objektverfolgung: Physiologie

Vor der bereits erwähnten Arbeit von Dodge und Cline hatten bereits im letzten Jahrhundert Donders [1847] und von Helmholtz [1866/1962] Augenbewegungen mit Hilfe von Nachbildern untersucht, allerdings ohne auf den Unterschied zwischen Sakkaden und Folgebewegungen zu stoßen. Den Beginn der quantitativen Modellierung von Augenbewegungen in der Sprache der linearen Systemtheorie markieren die Arbeiten von Westheimer [1954b,a]. Subsysteme wie sakkadenauslösende Neurone, okulare Motorneurone, Augenmuskeln etc. werden dabei durch ihre Übertragungsfunktion beschrieben; bei Bedarf wurde die Darstellung durch nichtlineare Glieder (z.B. Schwellenelemente) ergänzt. Die meisten Modelle dieses Typs postulieren eine serielle Signalkette vom Reiz bis zur Sakkade. Der systemtheoretische Ansatz wurde über Jahrzehnte hinweg intensiv verfolgt; stellvertretend seien die Arbeiten von Robinson und Fuchs [1969], Robinson [1972] und Becker und Jürgens [1979] genannt. Robinson und Fuchs konnten zeigen, daß sich durch elektrische Mikrostimulation im Superior Colliculus sowie den frontalen Augenfeldern von wachen Affen Sakkaden auslösen lassen, deren Richtung und Betrag vom genauen Reizort im jeweiligen Hirnareal abhängt. Damit war die Idee einer weitgehend retinotopen Abbildung des Sehraums im *Superior Colliculus* (auf der auch das von mir verwendete Modell beruht) experimentell abgesichert. Die Arbeit von Becker und Jürgens [1979] präsentiert ein Modell, das die außerordentlich kurzen Antwortzeiten bei Doppelsakkaden erklärt, indem es eine parallele Vorbereitung von zwei aufeinanderfolgenden Sakkaden einführt; damit weicht es als erstes quantitatives Modell vom Postulat einer seriellen Verarbeitung ab. Becker und Jürgens fordern denn auch als Konsequenz, zukünftige Modelle stärker an der offenkundig parallel organisierten Arbeitsweise des Nervensystems zu orientieren – eine Anregung, die von den zu dieser Zeit erstmals entstehenden Netzwerkmodellen umgesetzt wurde. Wichtige Anregungen zur Konstruktion von biologienahen Modellen gaben die Berichte über wesentlich verkürzte sakkadische Reaktionszeiten, wenn im Experiment der Fixationspunkt schon einige Zehntelsekunden vor dem Erscheinen des eigentlichen Blickziels weggenommen wird, bis hin zu extrem schnellen *Express-Sakkaden* mit Latenzen von unter 80 ms [Saslow, 1967; Fischer und Ramsperger, 1984]. Dieser ausgeprägte *Gap Effect* wurde als Hinweis auf zwei sich gegenseitig inhibierende Teilsysteme für Sakkaden und Fixation aufgefaßt; diese Sichtweise konnten spätere Arbeiten präzisieren und untermauern [Munoz und Wurtz, 1992, 1993a,b, 1995a,b].

Die Grundlage für das in dieser Arbeit verwendete neuronale Netz bildet der Vorschlag von Amari [1977], neuronale Felder mit lateraler Inhibition zur Simulation von Aufmerksamkeitsprozessen zu verwenden (vgl. Kap. 3.5). Während Amari sich in erster Linie auf die analytische Behandlung der Dynamik des von ihm vorgeschlagenen neuronalen Feldes konzentriert, wurde das Modell 1995 von Kopecz aufgegriffen und – mit einigen Modi-

fikationen – zur quantitativen Modellierung des *Gap Effect* verwendet [Kopecz et al., 1995; Kopecz und Schöner, 1995]. Von besonderem Interesse ist dabei nicht allein die quantitative Beschreibung dieser Effekte, sondern ebenso die gleichzeitige gute Übereinstimmung der Verschaltungsstruktur des Modells mit den bekannten neurobiologischen Befunden über die Steuerung von Augenbewegungen (vgl. auch Kap. 2.3.3).

1.4.2 Objektverfolgung in der technischen Bildverarbeitung

Die Entwicklung technischer Bildanalysesysteme begann im industriellen Fabrikationsbereich. Dort sind eingeschränkte Aufgabenstellungen und kontrollierte Bedingungen der Regelfall; dafür ist die Genauigkeit, mit der die Aufgabe gelöst wird, entscheidend. Als Beispiel mag ein Roboter dienen, der ein Werkstück auf einem Förderband auffinden, identifizieren und anschließend an einer bestimmten Stelle anbohren soll. Eine grobe Lokalisation ist unproblematisch: Bei bekannter, aber unterschiedlicher Farbe von Förderband und Werkstück liefert eine einfache Farbsegmentierung die Zugehörigkeit der Pixel zum Objekt bzw. Hintergrund (Einzelheiten zur Segmentierung s.u.). Um in einer solchen Umgebung die genaue Lage von Werkstücken zu bestimmen und sie ggf. zu verfolgen, kommt typischerweise ein *modellbasiertes* Verfahren zum Einsatz: Ein 2D- oder 3D-Modell des Werkstücks wird unter variablem Winkel auf die Ebene des Förderbandes projiziert. Der Bildanalyse-Algorithmus versucht dann die wenigen freien Parameter (Projektionswinkel) des Modells so anzupassen, daß eine möglichst gute Übereinstimmung zwischen modelliertem und echtem Bild entsteht. Diese Vorgehensweise wird als *Anpassungs-* oder *Matching-Verfahren* bezeichnet, die dabei verwendete Übereinstimmungs- bzw. Fehlerfunktion als *Matching-Funktion*.

Mit einer geeigneten Vorverarbeitung lassen sich derartige Verfahren auch erfolgreich auf reale Szenen anwenden; als Beispiel mag die folgende Analyse der Szene 'Durlacher Tor' (vgl. Kap. 5) sowie weiterer vergleichbarer Sequenzen durch Nagel [1985], Koller et al. [1993] und Graefe [1995] dienen. Abb. 1.1 zeigt eine Anpassung verschiedener 3D-Fahrzeugmodelle an die Szene.

Die Wirksamkeit solcher Verfahren beruht auf der Tatsache, daß erstens nur eine begrenzte Anzahl von Objekttypen relevant für die jeweilige Problemstellung ist und zweitens das Erscheinungsbild dieser Objekte im Prinzip bekannt ist. Ein prinzipieller Nachteil liegt darin, daß für jeden Objekttyp ein eigenes Modell notwendig ist – eine Verfolgung ohne Objektwissen ist nicht möglich.

Das in dieser Arbeit verwendete neuronale Netz verfolgt den umgekehrten Ansatz: Auf eine exakte Verfolgung von Objekten durch Einsatz von spezifischem Objektwissen wird verzichtet; dafür ist das verwendete System allgemein einsetzbar und kann jedes Objekt verfolgen, das sich in einer Szene durch seine Bewegung vom Hintergrund abhebt. Dies entspricht dem Gestaltgesetz vom *gemeinsamen Schicksal* [Wertheimer, 1912].

1.4.3 Szenensegmentierung

Wie das o.g. Beispiel mit Förderband und Werkstück zeigt, sind Segmentierung einer Szene und die Identifikation bzw. Verfolgung einzelner Objekte darin eng miteinander ver-

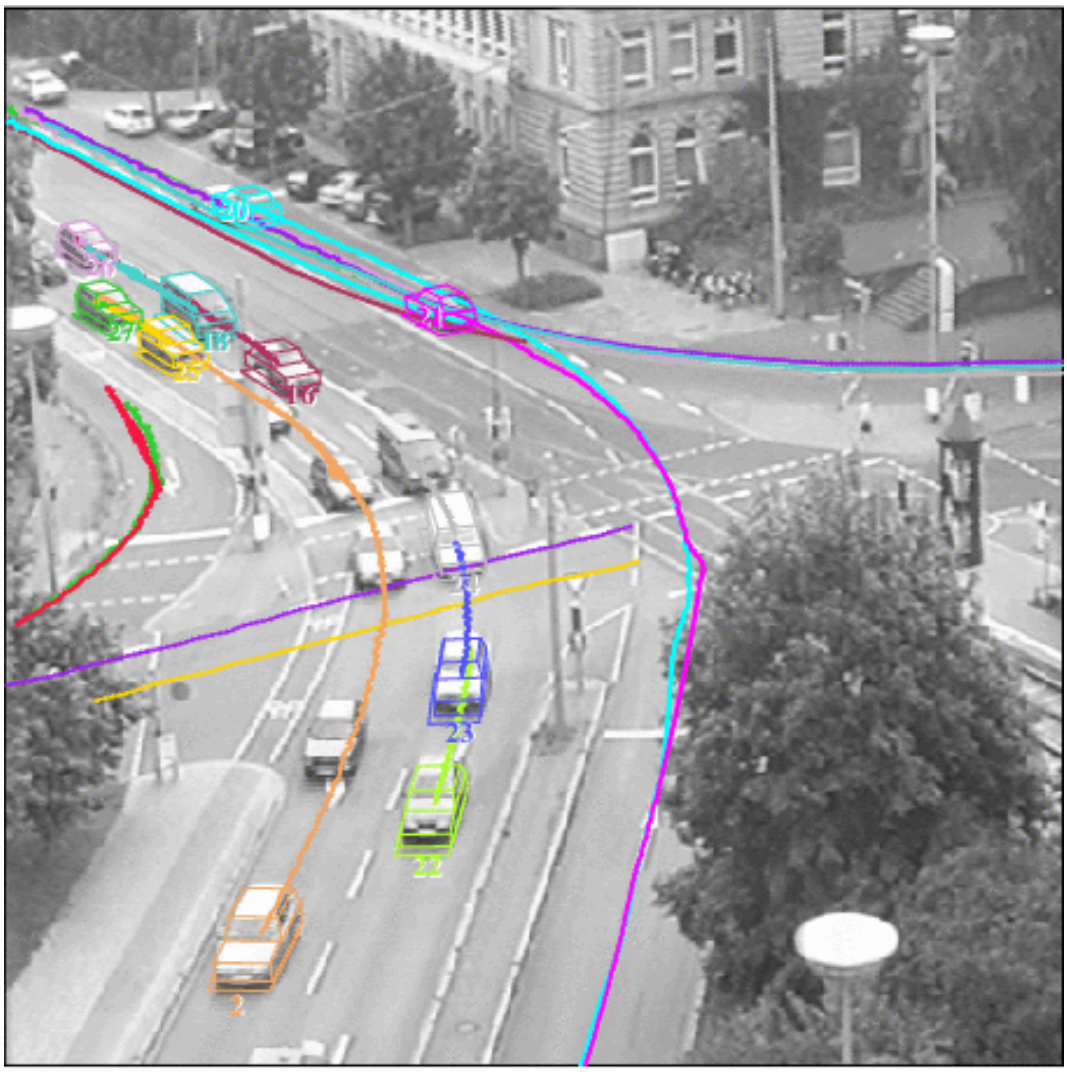


Abbildung 1.1: Anwendung eines Matching-Verfahrens auf eine Verkehrsszene. Das 3D-Drahtgittermodell eines PKW wird unter unterschiedlichen Winkeln in das Bild projiziert und die jeweilige Abweichung bestimmt; die eingezeichneten Kantenbilder markieren die Minima der Fehlerfunktion. Die Trajektorien wurden durch fortlaufende Anpassung der Modellparameter bestimmt. (Aus: [Leuck und Nagel, 1999])

wandte Probleme: Ohne eine Vorsegmentierung ist das Auffinden bzw. die Identifikation von Objekten kaum möglich; andererseits ist die Segmentierung interessierender Objekte zumindest bei bewegten Szenen nur schwer ohne ein geeignetes Verfolgungssystem zu realisieren; dieses sollte idealerweise auch schon vor der endgültigen Modellanpassung in der Lage sein, eine Verfolgung zu starten.

Die genannte Farbsegmentierung ist ein typisches Beispiel, wie man unter speziellen Umständen mit einfachen Algorithmen eine Szene in Objekt(e) und Hintergrund unterteilen kann. Diese einfachen Verfahren müssen aber überall da versagen, wo die Anhaltspunkte zur Unterteilung in komplexen Merkmalskombinationen ‘verborgen’ sind, etwa bei einer belebten Straßenszene mit unterschiedlichen Fahrzeugen und Fußgängern.

An diesem Punkt kommt der neurobiologische Ansatz zum Zug: Da menschliche Be-

obachter derartige Szenen schnell und sicher analysieren können, liegt es nahe, bekannte oder hypothetische Mechanismen des menschlichen Sehsystems in technische Systeme zu integrieren. Umgekehrt läßt sich aus dem Ergebnis solcher Integrationsversuche bis zu einem gewissen Grad ersehen, welche der vermuteten Mechanismen im Gehirn tatsächlich zu brauchbaren Ergebnissen bei der Verarbeitung natürlicher Bilder führen.

Zunächst scheinen die beiden Ansätze sehr gegensätzlicher Natur zu sein: Es erscheint z.B. nicht sinnvoll, eine einfache Farbsegmentierung mit neuronalen Netzen zu implementieren. Bezugspunkte ergeben sich dort, wo Operatoren zur Extraktion von Bildmerkmalen zunächst lokal angewandt und anschließend mit ihrer Nachbarschaft verknüpft werden. Ein retinotop über die Eingangsszene gelegtes neuronales Netz (z.B. zur Kantenextraktion) stellt nämlich nichts anderes als eine lokale Rechenvorschrift dar; die Eingangsfunktion visueller Modellneurone wird in Anlehnung an den biologischen Sprachgebrauch als *rezeptives Feld* bezeichnet. Die Anweisungen für die weitere Verrechnung mit der Nachbarschaft sind implizit in den Gewichtungsfaktoren der Verbindungen enthalten, die entsprechend als *synaptische Gewichte* bezeichnet werden. Ein prominentes Beispiel für eine solches Segmentierungsverfahren ist das *Region Growing* [Ballard und Brown, 1982; Gonzalez und Woods, 1992], bei dem von einem gleichmäßigen Raster von Startpunkten aus benachbarte Bildbereiche mit ähnlichen Charakteristika zu Regionen zusammengefaßt werden. Nach einem ähnlichen Grundgedanken versucht die *Hough-Transformation*, vorgegebene Muster (z.B. Geradenstücke) zunächst aufzufinden und dann soweit wie möglich in der Nachbarschaft wiederzuentdecken [Hough, 1962]. Das *Region Growing* ist inhaltlich eng mit der Linking-Idee von Eckhorn et al. [1990] verwandt (vgl. Kap. 3.3): Beide Ansätze implementieren *Bottom-Up* eine Ähnlichkeitsprüfung für benachbarte Bildorte, kombiniert mit einer Schwellenoperation. Allerdings werden beim *Region Growing* die Farbwerte der Pixel direkt verwendet, d.h. es findet keine Vorverarbeitung durch rezeptive Felder statt. Dies hat zur Folge, daß das Verfahren nur in relativ einfachen Situationen wie dem oben beschriebenen Roboterbeispiel eine befriedigende Segmentierung liefert.

Eckhorn et al. erreichten mit der Linking-Architektur u.a. eine **robuste Signalverarbeitung** durch neuronale Merkmalsdetektoren: Neuronen, die das gleiche Merkmal an benachbarten Orten codieren, unterstützen sich gegenseitig in ihrer Aktivität. Damit werden die einzelnen Detektoren robuster gegenüber Störungen wie sie z.B. durch schlechte Sichtbedingungen, teilweise Verdeckung oder internem (Membran-) Rauschen entstehen können. In Anlehnung an biologisch nachgewiesene *modulatorische Synapsen* führten Eckhorn et al. dabei als zuätzliche Besonderheit eine *multiplikativ* wirkende Nachbarschaftskopplung ein. Die diese Kopplung vermittelnden Verbindungen bezeichneten sie als *Linking-Synapsen*. Die Implikationen von additiver bzw. multiplikativer Nachbarschaftskopplung für das generelle Antwortverhalten der Neurone werden in Kap. 3.3.5 analysiert; Kap. 6 erweitert diese Überlegungen um ein stochastisches Modell zur Konturdetektion bei verschiedenen Störungen.

Ebenfalls aus der Neurobiologie stammt das in dieser Arbeit verwendete Konzept, die Zeit als Codierungsdimension zu verwenden: Die Zusammengehörigkeit von Bildelementen wird dabei im Grad der synchronen Aktivität der den jeweiligen Bildelementen zugeordneten Neurone codiert; desynchronisierte Aktivität bedeutet entsprechend 'keine Zusammengehörigkeit'. Diese *Synchronisationshypothese* wurde unabhängig von mehreren Au-

toren vorgeschlagen, zunächst als theoretischer Ansatz zur Lösung des *Binding-Problems* [Schneider et al., 1983; von der Malsburg und Schneider, 1986; Eckhorn et al., 1990]. Von experimenteller Seite bekam die Synchronisationshypothese starke Unterstützung, als im Cortex von Katzen und Affen Oszillationen entdeckt wurden, die im Frequenzbereich von 40–90 Hz liegen und an den visuellen Reiz gekoppelt sind [Eckhorn et al., 1988; Gray und Singer, 1989; Gray et al., 1990; Eckhorn et al., 1993]. Die Entstehung und Dynamik eines einfachen Typs solcher Oszillationen durch das Wechselspiel von neuronaler Exzitation und Inhibition werden in Kap. 4 behandelt. Dabei wird im Modell die Trennung von Objekten durch eine gegenphasige Aktivität codiert; neuere Ergebnisse deuten allerdings darauf hin, daß die tatsächliche Dynamik im Gehirn noch weit komplexer ist und eine explizit gegenphasige Aktivität getrennter Neurone nicht auftritt [Gail et al., 1999a; Gabriel und Eckhorn, 1999].

2 Wissenschaftliche Grundlagen

2.1 Neurobiologische Grundlagen

Das Nervensystem hat im Organismus die Aufgabe, sensorische Informationen aus der Umwelt zu verarbeiten, in geeigneter Weise zu verknüpfen und daraus adäquates Verhalten zu generieren (d.h. ein Verhalten, das dem Organismus das Überleben ermöglicht).

Die Fortleitung der Information erfolgt dabei in gerichteter Weise über die Nervenbahnen, die ihrerseits aus *Nervenzellen* oder *Neuronen* aufgebaut sind. Man unterscheidet *afferente* Nervenbahnen, die von den Sinnesorganen zum Gehirn ziehen, und *efferente* Bahnen, die in umgekehrter Richtung verlaufen. Fast alle Organe im menschlichen Körper verfügen über Fasern beider Typen, d.h. ein Informationsaustausch ist in beiden Richtungen möglich. Ebenso sind praktisch alle Verbindungen zwischen Gehirnarealen reziprok vorhanden.

Die kleinste funktionelle Einheit des Nervensystems bildet die Nervenzelle. Dieser Zelltyp ist in Aufbau und Funktion genau auf die Aufgabe der Informationsverknüpfung und -weiterleitung zugeschnitten; seine Funktion wird im folgenden Abschnitt kurz vorgestellt.

2.2 Aufbau einer Nervenzelle

Der Aufbau von Nervenzellen unterscheidet sich nicht grundlegend von dem anderer Zelltypen: Die Zelle ist durch eine elektrisch nichtleitende und für Ionen weitgehende undurchlässige Membran von der Umgebung abgegrenzt. Dadurch können sich über der Membran sowohl elektrische Spannungs- als auch Konzentrationsgefälle aufbauen. Aktive Pumpmechanismen halten außerhalb der Nervenzelle ständig einen Na^+-Überschuß aufrecht; innerhalb herrscht ein K^+-Überschuß. Cl^--Ionen können die Membran passieren; im elektrochemischen Gleichgewicht stellt sich dadurch eine Spannung von ca. $-70\ mV$ gegenüber dem Zelläußeren ein. Diese pflanzt sich entlang der isolierenden Membran fort, allerdings nimmt sie mit der Entfernung vom Entstehungsort ab.

Diese Eigenschaften lassen sich in vereinfachter Form im Ersatzschaltbild aus Abb. 2.2 darstellen: Jedes Stück einer Zellmembran verhält sich wie eine elektrische Kapazität mit parallel geschalteter Restleitfähigkeit. Die lokalen RC-Glieder bilden in ihrer Gesamtheit ein *verlustbehaftetes elektrisches Kabel*, auf dem sich Potentialdifferenzen in Längsrichtung fortpflanzen können. Lokal verhält sich jedes 'Kabelstück' wie ein Tiefpaß erster Ordnung.

Eine wesentliche anatomische Eigenschaft von Nervenzellen ist die Ausbildung langer, vom Soma ausgehender röhrenförmiger Fortsätze, die als *Nervenfasern* bezeichnet werden. Aufgrund der o.g. Eigenschaften können sie elektrische Potentialunterschiede in ihrer

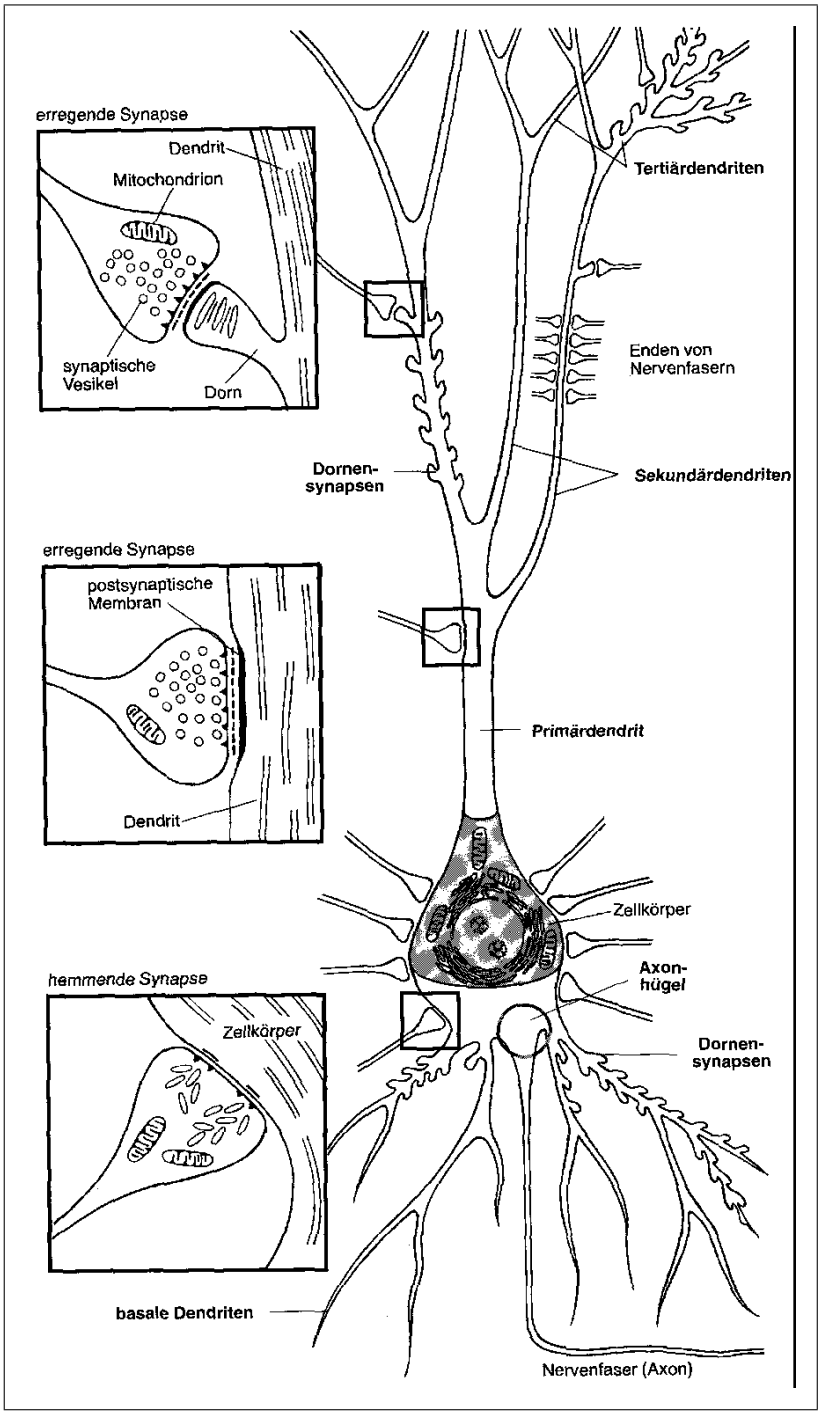


Abbildung 2.1: Darstellung einer Nervenzelle mit Dendriten, Soma und Axon. Während die Dendriten von anderen Zellen durch verschiedene Synapsentypen vermittelte Erregung aufsummieren und zum Soma hinleiten, entscheidet die Erregung am Soma (insbesondere am Axonhügel) über die Auslösung eines Aktionspotentials. Dieses wird ggf. durch das Axon weitergeleitet und wiederum über Synapsen an andere Zellen weitergegeben. (Nach: Roth und Prinz [1996])

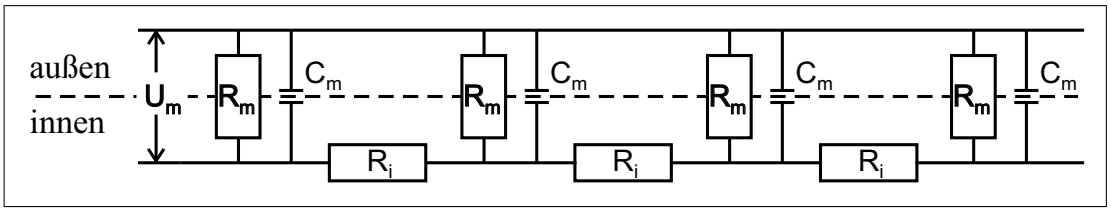


Abbildung 2.2: Ersatzschaltbild der Zellmembran einer Nervenfaser. Die elektrische Leitfähigkeit parallel zur Membran ist wesentlich größer als diejenige über die Membran hinweg. Lokal wirkt jedes Stück Membran wie ein Kondensator mit parallel geschaltetem Transmembranwiderstand.

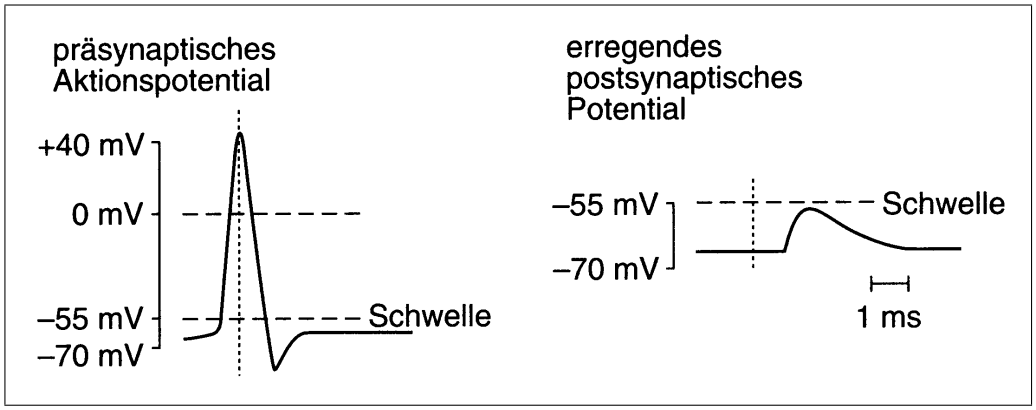


Abbildung 2.3: Typischer Zeitverlauf eines Aktionspotentials und des davon ausgelösten exzitatorischen postsynaptischen Potentials (EPSP). Die Verzögerung und Verbreiterung gegenüber dem auslösenden Aktionspotential kommt i.w. durch die Zeitspanne zustande, die für die präsynaptische Transmitterfreisetzung und die Diffusion durch den synaptischen Spalt benötigt wird. (Aus: [Roth und Prinz, 1996])

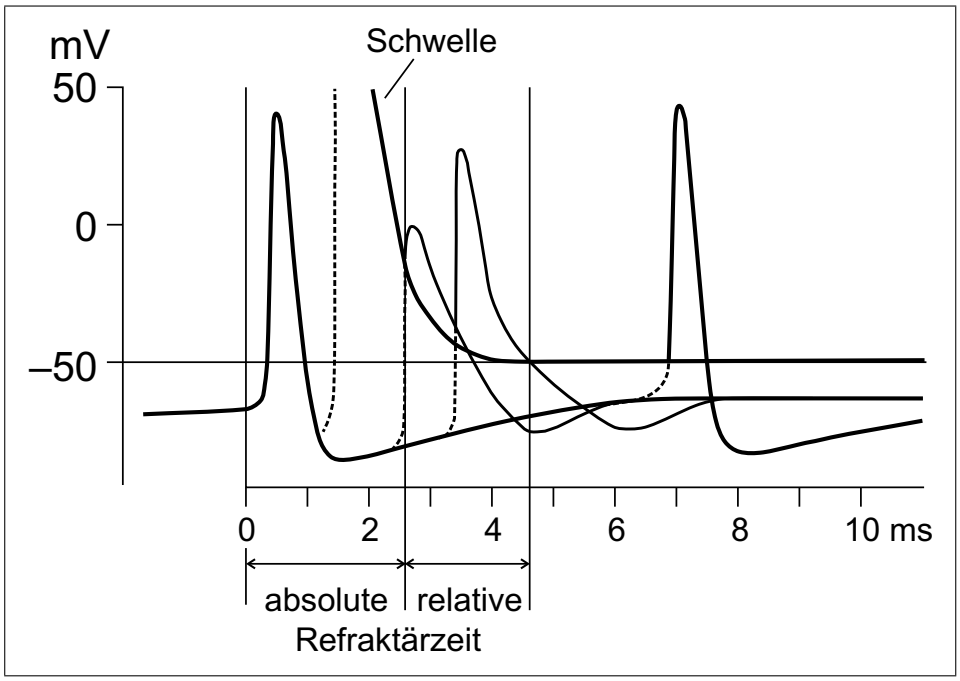


Abbildung 2.4: Absolute und relative Refraktärphase der Nervenzelle nach einem Aktionspotential. Die Schwelle nähert sich während der relativen Refraktärzeit kontinuierlich wieder ihrem Ruhewert an. (Aus: [Schmidt und Thews, 1996])

Längsrichtung weiterleiten, wobei diese Reizleitung in beiden Richtungen (zum und vom Soma) erfolgen kann.

Übersteigt das Membranpotential der Zelle allerdings einen bestimmten Wert (ca. $-35\ mV$), so wird neben dem bisher beschriebenen *passiven* Leitungsmechanismus der sogenannte *Hodgkin-Huxley-Zyklus* in Gang gesetzt [Hodgkin und Huxley, 1952]: Durch die sich öffnende Na^+-Kanäle strömen zunächst positive Na^+-Ionen in die Zelle ein, das Membranpotential steigt bis auf ca. $+30\ mV$an. Kurze Zeit später öffnen sich auch die K^+-Kanäle, was einen Einstrom von K^+-Ionen zur Folge hat; das Membranpotential sinkt wieder bis in die Nähe des Ruhewertes ab. Abb. 2.3 zeigt den zugehörigen Zeitverlauf des Membranpotentials; der gesamte Prozeß wird als *Aktionspotential* oder auch *Spike* bezeichnet. Er läuft, einmal angestoßen, nach einem stereotypen Muster ab. Das gesamte Aktionspotential ist normalerweise nach ca. 2 ms beendet; danach sind die Na^+-Kanäle zwar geschlossen, aber nicht im Ruhezustand. In diesen kehren sie erst allmählich zurück, was eine erneute Auslösung von Aktionspotentialen zunächst unmöglich macht und während der Übergangsphase von etwa 50 ms erschwert. Diese Zeitspannen werden als *absolute* bzw. *relative Refraktärzeit* bezeichnet und sind in Abb. 2.4 dargestellt.

Durch die kurzzeitig positiven Spannungswerte der Zellmembran werden auch angrenzende Bereiche der Membran depolarisiert, so daß auch hier lokale Aktionspotentiale ausgelöst werden und so eine *aktive Fortleitung* des Spikes entlang einer Nervenfaser möglich ist. Dies geschieht typischerweise ein einem speziellen Fortsatz des Neurons, dem *Axon*. Die Leitungsgeschwindigkeiten betragen hier 1–100 m/s. Die anderen, *Dendriten* genannten Fortsätze der Zelle leiten dagegen auf passive Weise elektrische Signale zum Zellkörper hin.

2.2.1 Synapsen

Das Axon endet in einem Büschel sogenannter *Synapsen*, die in unmittelbarer Nähe der Dendriten anderer Neurone liegen; ein eintreffendes Aktionspotential bewirkt hier die Ausschüttung eines sogenannten *Neurotransmitters*, der an der dendritischen Zellmembran spezielle, für diesen Transmitter empfindliche Ionenkanäle öffnen kann. Auf diese Weise kann die elektrische Erregung nach der räumlichen Weiterleitung im Axon an andere Zellen weitergegeben werden (*exzitatorische Synapse*); ebenso gibt es *inhibitorische Synapsen*, deren Neurotransmitter Erregungsprozesse am folgenden Neuron erschweren.

2.3 Das visuelle System des Menschen

2.3.1 Retina und Sehbahn

Der Gesichtssinn stellt für einen (gesunden) Menschen den mit Abstand informationsreichsten Sinneskanal dar. So können z.B. visuelle Eindrücke widersprüchliche Informationen aus anderen Sinnesmodalitäten, etwa vom Gleichgewichtssinn, leicht überwiegen; umgekehrt ist dies sehr viel seltener der Fall. Insbesondere Informationen über die Beschaffenheit, Lage und Anordnung der uns umgebenden Gegenstände werden vom Sehsystem präzise an andere Gehirnfunktionen wie die Greifmotorik übermittelt.

Allerdings sind diese Informationen zunächst nur indirekt in Form einer zweidimensionalen Helligkeits- bzw. Farbverteilung auf der Netzhaut des Auges verfügbar. Aus dieser indirekten, aber räumlich und zeitlich hochaufgelösten Informationsquelle ein zuverlässiges Bild der Umgebung zu gewinnen, ist die Aufgabe des visuellen Systems. Abb. 2.5 zeigt die neurale Sehbahn in der Übersicht. In der Netzhaut (*Retina*) des Auges werden die Lichtsignale in Folgen von Aktionspotentialen umgesetzt, die vom Sehnerv zunächst zum seitlichen Kniehöcker (*Corpus Geniculatum Laterale*, CGL) und von dort zum visuellen Cortex weitergeleitet werden. Bereits vor dem CGL treffen sich die Sehnerven beider Augen am *Chimasma Opticum* und überkreuzen sich teilweise. Beide CGL erhalten also Signale vom jeweils gleichseitigen (ipsilateralen) und gegenüberliegenden (contralateralen) Auge. Von den Relaiszellen des CGL ziehen weitere Fasern als *Sehstrahlung* zum primären visuellen Areal V1 des Cortex. Dabei wird besonders der mittlere (nasale) Teil des Gesichtsfeldes auf die contralaterale Hirnhemisphäre abgebildet. Im visuellen Cortex wird der Input auf zahlreiche Gebiete mit spezialisierten Eigenschaften aufgeteilt. Zu bemerken ist, daß praktisch alle Verbindungsstrukturen im visuellen System reziprok, d.h. hin- und rücklaufend vorhanden sind, mit Ausnahme derjenigen von der Retina zum CGL.

Die Retina verfügt über ca. 10^8 Rezeptorzellen mit unterschiedlichen Eigenschaften. Die wichtigste Einteilung ist die zwischen *Stäbchen* und *Zapfen*: Während erstere auf das Sehen bei Nacht angepaßt sind und eine hohe Empfindlichkeit aufweisen, werden letztere erst bei stärkerer Beleuchtung aktiv und vermitteln dafür Farbensehen und gute räumliche Auflösung.

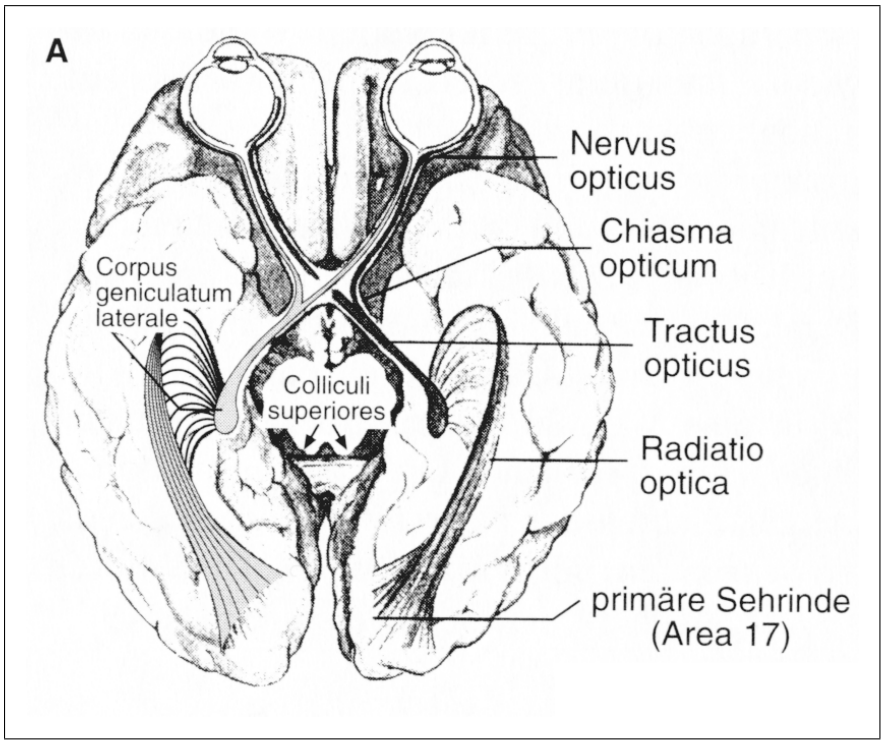


Abbildung 2.5: Schematische Darstellung der menschlichen Sehbahn. Die von der Retina kommenden Nervenfasern werden (teils überkreuzt) über die beiden CGL zum visuellen Cortex und zu den Colliculi Superiores weitergeführt. (Aus: [Roth und Prinz, 1996])

2.3.2 Rezeptive Felder

Während die Rezeptorzellen selbst generell auf Lichteinfall reagieren, stellt man fest, daß bereits die nachgeschalteten retinalen Ganglienzellen, die Aktionspotentiale für den Sehnerv generieren, bevorzugt auf *konzentrische Hell-Dunkel-Kontraste* ansprechen, also nicht durch großflächige homogene Lichtreize angeregt werden. Man unterscheidet On-Zentrum- und Off-Zentrum-Neurone, je nachdem ob ein heller Lichtpunkt auf dunklem Grund oder die umgekehrte Konfiguration den geeigneten Reiz für das jeweilige Neuron darstellt. Beide Typen treten etwa gleich häufig auf. Ein solches lokalisiertes räumliches Empfindlichkeitsprofil bezeichnet man als *rezeptives Feld* (RF) der Zelle. Während sich im CGL ähnliche RFs wie in der Retina finden, treten bereits in V1 kompliziertere Formen wie z.B. längliche, orientierte RF-Profile auf. In den höheren Schichten werden die RF-Eigenschaften bei wachsender RF-Größe zunehmend komplexer.

Typisch ist jedoch, daß ähnliche Merkmale häufig von benachbarten Neuronen codiert werden, sowohl was den Ort im Sehraum als auch beispielsweise die Vorzugsrichtung (Orientierung) eines RFs betrifft. Diese Anordung wird als *retinotope Organisation* bezeichnet, da Nachbarschaftsbeziehungen aus der Retina in hohem Umfang erhalten bleiben.

In der Retina bzw. dem CGL findet bereits eine *Vorverarbeitung* der einfallenden Lichtreize statt, die eine starke Informationsreduktion zur Folge hat, da Kontrastkanten in natürlichen Bildern wesentlich seltener als homogene Flächen sind (s. auch Kap. 6). Den ca. 10^8 Rezeptoren der Retina stehen nur ca. 10^6 Fasern im Sehnerv gegenüber.

Neben der räumlichen Vorverarbeitung findet aber auch eine besondere Umsetzung der zeitlichen Reizeigenschaften statt. Man unterscheidet hierbei drei Typen von Ganglienzellen: Die *Magno-* oder Y-Zellen reagieren auf Lichteinfall in ihrem RF mit einer schnellen, aber vorübergehenden (transienten) Aktivität. Die Antwort der *Parvo-* oder X-Zellen beginnt später, hält dafür aber wesentlich länger an. Schließlich gibt es noch die *Konio-* oder W-Zellen, die weitverzweigte Dendritenbäume und große Antwortlatenzen aufweisen und häufig bewegungsempfindlich sind. Diese Unterteilung in magno-, parvo- und koniozelluläres System findet sich auch im CGL und den Projektionen zu den verschiedenen kortikalen Arealen wieder: Das magnozelluläre System speist in erster Linie Areale, die mit der Verarbeitung schneller, bewegter Reize befaßt sind. Demgegenüber projiziert das parvozelluläre System hauptsächlich zu den Arealen, die für Form- und Figurwahrnehmung zuständig sind.

2.3.3 Hirnstrukturen innerhalb des visuellen Systems

Abb. 2.6 zeigt eine Übersicht über die wichtigsten visuellen Areale des Cortex. Die in V1 einlaufenden Signale werden nach V2 und V3 weiterverteilt. Während die Zellen in Area V4 besonders empfindlich für Farb- und Texturreize sind, spielen MT, MST und FST eine wichtige Rolle bei der Bewegungsverarbeitung. Diese Aufgabenteilung wird bereits durch die oben besprochene Einteilung in magno- und parvozelluläres System vorbereitet.

Für die Steuerung der Augenbewegungen (insbesondere die Auslösung von Sakkaden) sind die – entwicklungsgeschichtlich viel älteren – *Colliculi Superiores* verantwortlich. Sie erhalten sowohl direkten visuellen Input von der Retina (am CGL vorbei) als auch von anderen Sinnesmodalitäten (akustisch, somatosensorisch) und verfügen über viele bi- und

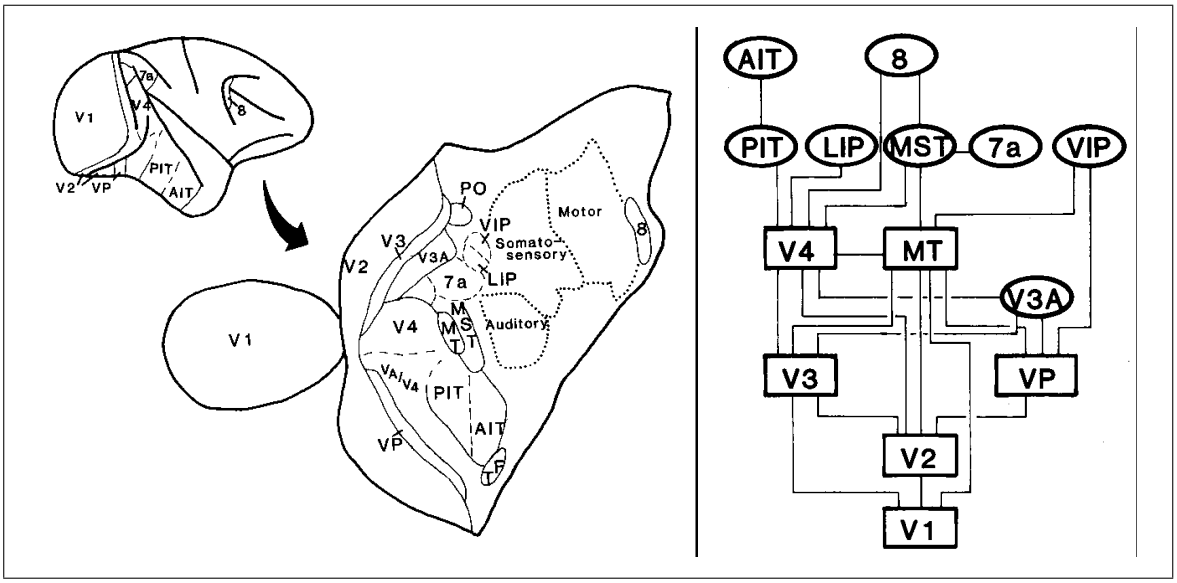


Abbildung 2.6: Übersicht über die wichtigsten visuellen Cortexareale und ihre vermuteten Verschaltungen. Die mittlere Abbildung zeigt eine 'aufgeklappte' Cortexoberfläche. Die Abkürzungen bedeuten: V1–V4A visuelle Areale (s. Text), AIT anterior inferotemporal, LIT lateral intraparietal, MST medial superior temporal, MT medial temporal, PIT posterior inferotemporal, PO parieto-occipital, VIP ventral intra parietal, VP ventral posterior. (Aus: [van Essen, 1987])

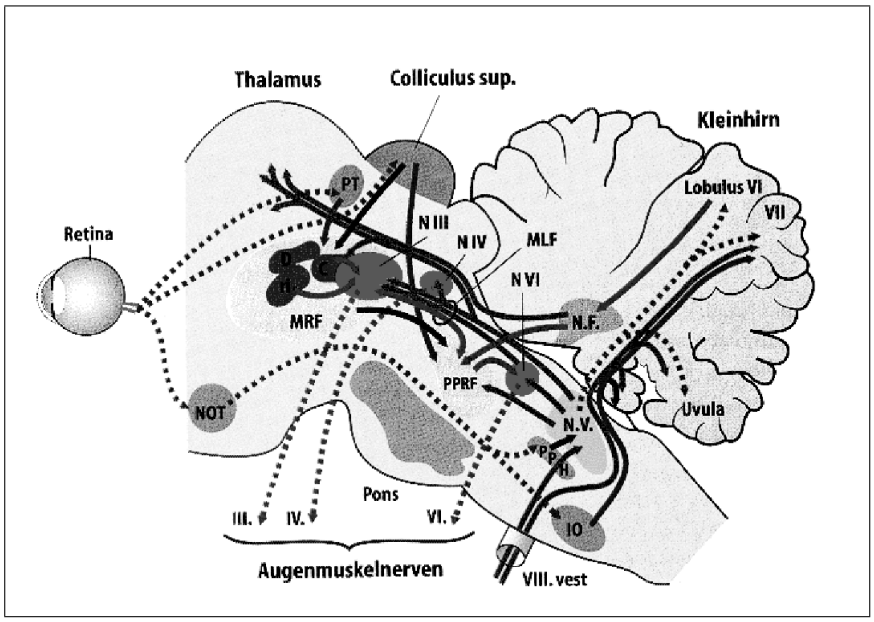


Abbildung 2.7: Übersicht über die an der Steuerung von Augenbewegungen beteiligten Hirngebiete. Bewegungsempfindliche Ganglienzellen (W-Zellen) der Retina projizieren sowohl direkt in die Colliculi Superiores als auch zum Kern des optischen Traktes (NOT) und in das Prätektum (PT). Während letzteres für die Steuerung von Vergenzbewegungen zuständig ist, werden in MRF und PPRF kortikale Signale integriert. Die eigentliche Steuerung der Augenmuskeln obliegt den Kernen N III, IV und V. (Aus: [Schmidt und Thews, 1996])

trimodale Neurone. Ihre Aufgabe liegt in erster Linie in der reflektorischen Steuerung von Blickbewegungen. Über den Pulvinar (eine Unterstruktur des Thalamus) stehen sie mit den kortikalen Bewegungsarealen (Area MT) in Wechselwirkung, so daß Blickbewegungen sowohl subkortikal als auch kortikal ausgelöst und beeinflußt werden können. Abb. 2.7 zeigt eine Übersicht über die vielfältigen Areale, die an der Steuerung von Blickbewegungen beteiligt sind.

Die sakkadenauslösenden Neurone im Colliculus Superior sind ähnlich wie die frühen visuellen Cortexareale als zweidmensionale retinope Karte organisiert, d.h. benachbarte Blickziele im Sehraum werden durch benachbarte Neuronengruppen codiert. Daneben existieren auch Neurone, die während der stabilen Fixation aktiv sind, also wenn gerade keine Augenbewegung durchgeführt werden soll. In Kap. 3.5.5 wird ein Modell behandelt, das diese scheinbar gegensätzlichen Aufgaben in einheitlicher Weise beschreibt und erfolgreich für die quantitative Vorhersage von sakkadischen Reaktionszeiten beim Menschen eingesetzt wurde [Kopecz, 1995].

2.4 Codierung im visuellen System

In Kap. 2.3 haben haben wir festgestellt, daß die zentrale Aufgabe des visuellen Systems in der Erzeugung einer *internen Repräsentation* der Umwelt besteht, die genügend Informationen enthält, um dem Organismus ein adäquates Verhalten zu ermöglichen. Diese Aufgabe bewältigt es offensichtlich durch die Bereitstellung einer Vielzahl unterschiedlicher Teilsysteme, die jeweils verschiedene Teilaufgaben 'übernehmen' und diese parallel bearbeiten. Dabei werden die komplexen Informationen aus der Umwelt zunächst gefiltert (vorverarbeitet) und anschließend in die Hierarchie der visuellen Cortexareale eingespeist. Wie in Kap. 2.3 erläutert, haben die einzelnen kortikalen Areale teilweise hochgradig spezialisierte Funktionen, wobei die wichtigste Unterteilung die in ein transientes Bewegungs- und ein stationäres Kontur-Form-System darstellt. Innerhalb dieser Systeme läßt sich mit fortschreitender Verarbeitung eine zunehmende Komplexität der codierten Merkmale feststellen. Während die Neurone in V1 vorwiegend auf orientierte Balkenreize antworten, sind in IT bereits Neurone zu finden, die auf komplexe lokale Eckenmuster ansprechen. Ebenso sind viele der orientierungssensitiven Neurone in V1 gleichzeitig bewegungsempfindlich (und damit wenig spezifisch), während in Area MT praktisch alle Neurone eine hohe Spezifität für eine bestimmte lokale Bewegungsrichtung aufweisen. In höheren Schichten nimmt die Komplexität der codierten Merkmale weiter zu, während gleichzeitig die Abhängigkeit vom Ort im Sehraum schwächer wird.

Aus diesen Erkenntissen läßt sich prinzipiell verstehen, wie eine komplexe Szene zunächst durch lokale Merkmalsextraktion in ihre Bestandteile 'zerlegt' wird, die in höheren Stufen wieder zu komplexeren Formen zusammengesetzt werden. Dabei ist jedoch völlig unklar, wie die einzelnen, getrennten (auch komplexen) Merkmale einander so zugeordnet werden, daß am Ende eine einheitliche, konsistente Wahrnehmung entsteht (die zudem mit der physikalischen Realität möglichst weitgehend übereinstimmen sollte). Dieses Problem der Zusammengehörigkeit von Teilobjekten bzw. Merkmalen wird in der Literatur vielfach als *Binding*-Problem bezeichnet.

Wie bereits in Kap. 1 erwähnt, wurde praktisch gleichzeitig von mehreren Autoren

unabhängig voneinander die Idee der *zeitlichen Codierung* von Zusammengehörigkeit ins Spiel gebracht. Damit ist folgendes gemeint: Während einzelne Neuronen (bzw. kleine Neuronengruppen) durch ihre Aktivität die Anwesenheit und Stärke eines bestimmten Merkmals (etwa einer lokalen Orientierung) im Bild codieren, gibt dies noch keinen Hinweis darauf, welche der detektierten lokalen Linienelemente zu einem Gegenstand oder auch nur Linienzug gehören sollen. Ist jedoch die Aktivität der Neuronen so strukturiert, daß zusammengehörige Neuronengruppen (etwa solche, die im Verlauf einer geschlossenen Linie) liegen, zeitlich korreliert feuern, so läßt sich das Zusammenbinden der Teilobjekte auf elegante Weise bewerkstelligen, ohne die sonstigen Codierungseigenschaften der Neurone zu beeinträchtigen. Ebenso kann die Trennung von verschiedenen Objekten durch dekorrelierte bzw. desynchronisierte Aktivität dargestellt werden. Wie wir in Kap. 4.1 sehen werden, können genau diejenigen exzitatorischen Nachbarschaftsverbindungen, die z.B. zur Unterstützung durchlaufender Linien (entsprechend den Gestaltgesetzen) dienen, auch zuverlässig die Synchronisation der Aktivität der betroffenen Zellgruppen bewirken. Die Desynchronisation von Neuronengruppen, die zu trennende Objekte repräsentieren, läßt sich entsprechend durch einen gemeinsam wirkenden inhibitorischen Mechanismus erreichen, der mit den überall im Cortex vorhandenen inhibitorischen Interneuronen identifiziert wird. Die Wirkungsweise dieser Mechanismen wird in Kap. 4.1 im einzelnen vorgestellt.

Berücksichtigt man die auf den Nerven vorhandenen Verzögerungen in realistischer Weise, so muß man den Begriff der Synchronisation im Sinne eines allgemeineren Kohärenzbegriffs erweitern, d.h. die direkte Gleichzeitigkeit durch eine feste, aber beliebige Phasenbeziehung ersetzen (die immer noch im Rahmen einer gekoppelten, oszillatorischen Aktivität zu definieren ist). Theoretische Untersuchungen, wie solche erweiterten Phasenkopplungen im Gesamtsignal noch festzustellen sind, wurden von Schanze und Eckhorn [1997] vorgestellt.

Den in dieser Arbeit verwendeten Modellen liegt die Synchronisationshypothese in ihrer ursprünglichen Form zugrunde, d.h. Objekte gelten als zusammengehörig, wenn die Aktivität der sie repräsentierenden Neurone in einem Zeitfenster von ca. 10 ms synchronisiert ist und als getrennt, wenn die zeitliche Verschiebung zwischen den Objekten sich um ungefähr eine Größenordnung davon unterscheidet.

2.5 Psychophysische Grundlagen

Aus den zahlreichen psychophysischen Phänomenen, die Rückschlüsse auf die Arbeitsweise des Sehsystems erlauben, sollen hier nur diejenigen vorgestellt werden, die für die vorliegende Arbeit unmittelbar von Bedeutung sind. Dies sind zum einen die bereits angesprochenen *Gestaltgesetze* (einschließlich Bewegung) und zum zweiten das Phänomen der visuellen Aufmerksamkeit in einer speziellen Ausprägung.

2.5.1 Die Gestaltgesetze

Die bereits in der Einleitung erwähnten Gestaltgesetze bilden eine Vielzahl empirischer Regeln, nach denen die menschliche Wahrnehmung Objekte als zusammengehörig, ‘sinn-

voll' oder auch ästethisch schön beurteilt. Die wichtigste dieser Regeln ist in unserem Zusammenhang diejenige vom *guten Verlauf* bzw. der *guten Form*; für bewegte Objekte tritt das Gesetz vom *gemeinsamen Schicksal* hinzu.

Wie der Leser am Beispiel in Abb. 2.8 (hoffentlich) selbst nachvollziehen kann, führt die Ansammlung einzelner gerader Linienstücke nicht notwendig zur Wahrnehmung einer Ansammlung unabhängiger Elemente. Statt dessen hat man den Eindruck zweier getrennter Teilstrukturen, die etwa dem Bild eines 'Wasserfalls vor einer Wand' entsprechen. Daraus läßt sich folgendes ersehen:

1. Kollinear angeordnete Linienstücke werden als zusammenengehörig wahrgenommen, wenn ihr Abstand nicht zu groß ist. Dies gilt auch, wenn die Stücke nicht exakt parallel sind.

2. Rechtwinkliges oder fast rechtwinkliges Aufeinandertreffen von Linien(-stücken) wirkt im Gegensatz dazu trennend. Im Zwischenbereich (Winkel ca. 30–60 Grad) ist die Wahrnehmung mehrdeutig. Hier hat die Umgebung einen starken Einfluß, d.h. Linienstücke, die bereits einem Linienzug zugeordnet sind, können nicht gleichzeitig zu einem anderen, kreuzenden gehören [Lüschow und Nothdurft, 1993].

Diese Feststellung wird als Gesetz vom *guten Verlauf* bezeichnet. In Kap. 6 ist genauer analysiert, inwiefern gerade diese Art des Zusammenbindens auf die statistischen Eigenschaften natürlicher Bilder abgestimmt ist.

Weitere derartige Gesetze betreffen die Wahrnehmung der Geschlossenheit von Konturen (*gute Form*) sowie die Anordnung von Bildelementen, insbesondere die Rolle von Ecken bei der Erzeugung von Scheinkonturen.

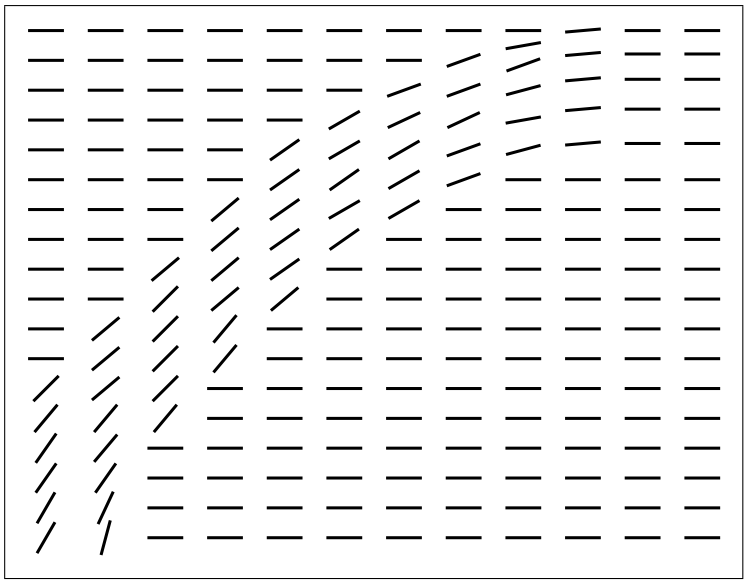


Abbildung 2.8: Beispiel zum Gesetz des guten Verlaufs. Annähernd kollinear angeordnete Liniensegmente erscheinen als zu einem Linienzug gehörig, auch wenn keine direkte physikalische Verbindung zwischen ihnen besteht.

2.5.2 Visuelle Aufmerksamkeit

Der Begriff *Aufmerksamkeit* oder *fokale Aufmerksamkeit* wird in der wissenschaftlichen Literatur in unterschiedlichen Bedeutungen verwendet. Im allgemeinen ist damit eine selektiv verbesserte und/oder beschleunigte visuelle Verarbeitungsleistung bei bestimmten Aufgaben gemeint. Man unterscheidet dabei zwischen unwillkürlicher und willkürlicher Aufmerksamkeit, je nachdem ob ein Ereignis die Aufmerksamkeit (und damit normalerweise auch den Blick) 'unwillkürlich' auf sich gezogen hat oder ob eine absichtliche Bevorzugung einer bestimmten Region im Blickfeld vorliegt. Letzteres geschieht häufig als Folge einer Anweisung der Form: 'Achte auf die rechte obere Ecke des Bildschirms!'. Beispiele für unwillkürliche Aufmerksamkeitsprozesse sind reflektorische Sakkaden zum Ort einer plötzlichen Bewegung bzw. Veränderung im Sichtfeld oder zu einem lauten Schallereignis. Daneben existieren noch Formen von Aufmerksamkeit, die nicht unmittelbar an den Ort im Sehraum gebunden sind, sondern andere Merkmalsdimensionen wie Farbe, Form, Bewegung und abstraktere Eigenschaften berücksichtigen (*nicht-fokale Aufmerksamkeit*).

Diese einfachen Beispiele zeigen bereits, daß es sich bei *Aufmerksamkeit* um ein komplexes Phänomen handelt, das sowohl Wahrnehmungs- als auch Handlungsaspekte umfaßt. Ich beschränke mich in dieser Arbeit auf die Modellierung von Phänomenen der *fokalen Aufmerksamkeit*, d.h. der Bezug zum Sehraum ist immer gegeben. Aufgabe des modellierten Aufmerksamkeitssystems ist es demnach, denjenigen Bereich einer Szene auszuwählen, der – je nach Aufgabenstellung – den wichtigsten oder prominentesten Input liefert und den Blick dorthin auszurichten.

Die unmittelbare Kopplung der Blickrichtung an die fokale Aufmerksamkeit ist eine starke Vereinfachung gegenüber der biologischen Situation; menschliche Versuchspersonen können den Aufmerksamkeitsfokus verlagern, ohne die Blickrichtung zu verändern. Das umgekehrte ist aber vermutlich nicht möglich; einer Sakkade geht immer eine Verlagerung der Aufmerksamkeit voraus. Andererseits stellt die Trennung von Aufmerksamkeit und Blickrichtung beim natürlichen Sehvorgang eine Ausnahme dar; normalerweise folgt die Blickrichtung innerhalb von Sekundenbruchteilen dem Aufmerksamkeitsfokus.

Generell ist die Selektivität in der sensorischen Verarbeitung im Organismus keineswegs auf das visuelle System beschränkt, sondern in ein globales Schema von Hin- und Wegwendungsreaktionen eingebettet, das dem Lebewesen in vielen Situationen eine zugleich schnelle und angemessene Reaktion auf äußere Reize ermöglicht. Räumlich gerichtetes Verhalten ist in einfacher Form bereits bei einfachen Lebewesen wie Bakterien als Chemo- oder Phototaxis zu beobachten. Bei komplexeren Lebewesen spricht man in diesem Zusammenhang von *Aufmerksamkeit*, wobei der Begriff wie erwähnt teilweise in sehr unterschiedlicher Bedeutung gebraucht wird, besonders was den Unterschied zwischen 'bewußter' und 'unbewußter' Aufmerksamkeit betrifft.

2.5.3 Mechanismen der Aufmerksamkeit

Durch welche neuronalen Mechanismen der Fokus letztendlich bevorzugt behandelt wird, ist unklar. Der einfachste Erklärungsansatz besteht in einer neuronalen Exzitation innerhalb und/oder einer Inhibition aller Bereiche außerhalb des Fokus. Diese Idee ist konform mit einer Reihe von Wahrnehmungsexperimenten, bei denen im Fokus erniedrigte, außer-

halb erhöhte Schwellen für die Kontrastwahrnehmung nachgewiesen wurden. Ähnliches gilt für die Wahrnehmung von Form und Bewegung, aber auch für komplexe Unterscheidungsaufgaben (Lesen).

Als natürliche Konsequenz aus den so veränderten Amplituden ergibt sich auch ein verändertes Zeitverhalten der betroffenen Neurone: Neurone, die durch Exzitation der vorgeschalteten Bereiche stärkeren Input erhalten, erreichen früher ihre Feuerschwelle, zeigen also auch früher eine erste Antwort auf einen neuen Reiz. Das Gegenteil gilt für gehemmte Neurone außerhalb des Fokus. Dabei ist für nachgeschaltete Neurone nicht zu unterscheiden, ob die Veränderung ihres Inputs durch eine veränderte Reizsituation in ihrem RF zustandekommt, oder 'nachträglich' durch einen Aufmerksamkeitseffekt verursacht wurde. Die Äquivalenz von physikalischer und aufmerksamkeitsbedingter Veränderung des neuronalen Inputs illustrieren eine Reihe von Arbeiten, die ich im folgenden Abschnitt kurz vorstelle. Dabei liegt der Schwerpunkt auf dem zeitlichen Aspekt der aufmerksamkeitsbedingten Veränderungen der Wahrnehmung; die beschriebenen Versuchssituationen geben keinen direkten Hinweis darauf, daß über eine anfängliche Beschleunigung/Verzögerung hinaus bestimmte Bereiche des Sehraums dauerhaft verstärkt bzw. unterdrückt werden (als Meßgröße fungieren allein die anfänglichen Latenzen). Dies mag damit zusammenhängen, daß alle beschriebenen Versuche mit einer einfachen, reflexiven Form von Aufmerksamkeit arbeiten, deren Wirkung möglicherweise nur kurz anhält – bewußte Konzentration auf einen bestimmten Bereich des Sehraums ist sehr wohl geeignet, dauerhafte Effekte zu erzeugen.

Diese Einschränkung auf reflektorische Aufmerksamkeitsmechanismen und die Betonung des zeitlichen Aspekts erscheint im Kontext des von mir verwendeten, funktional orientierten Modells sinnvoll: Ziel ist ja eine datengetriebene *Low-Level*-Segmentierung komplexer Bildinhalte in der Zeitdomäne, d.h. es stehen weder höhere (kortikale) Mechanismen zur Ausrichtung des Fokus zur Verfügung noch ist eine Veränderung von Amplituden zur Verbesserung der Segmentierung notwendig. Wie sich in Kap. 4.4 zeigen wird, kann sich eine relative zeitliche Dispersion der Bildbereiche dagegen positiv auf die Segmentierungsleistung des Systems auswirken.

2.5.4 Einige Experimente zum zeitlichen Aspekt von Aufmerksamkeit

Am besten wird das grundlegende Phänomen durch ein Experiment von Hikosaka et al. [1993a] verdeutlicht (Abb. 2.5.4): Die Versuchsperson fixiert ein kleines Kreuz in der Mitte eines weißen Bildschirms. Wird plötzlich ein schwarzer Balken eingeblendet, so wird er zwar als zeitlich transient, aber räumlich homogen, d.h. unverändert wahrgenommen. Wird aber einige Zehntelsekunden vor dem Balken als Hinweis ein kleines Quadrat links daneben für kurze Zeit eingeblendet ('geflasht'), so scheint der anschließend auftauchende Balken aus der so markierten Seite 'herauszuwachsen'; es entsteht eine scheinbare Bewegung nach rechts.

Die einfache Erklärung, daß die visuelle Verarbeitung in der Nähe des Hinweisreizes beschleunigt wird und so die Bewegungsillusion auslöst, hat sich in weiteren Experimenten bestätigt. In allen Fällen bringt eine entgegengesetzte physikalische Zeitdifferenz die

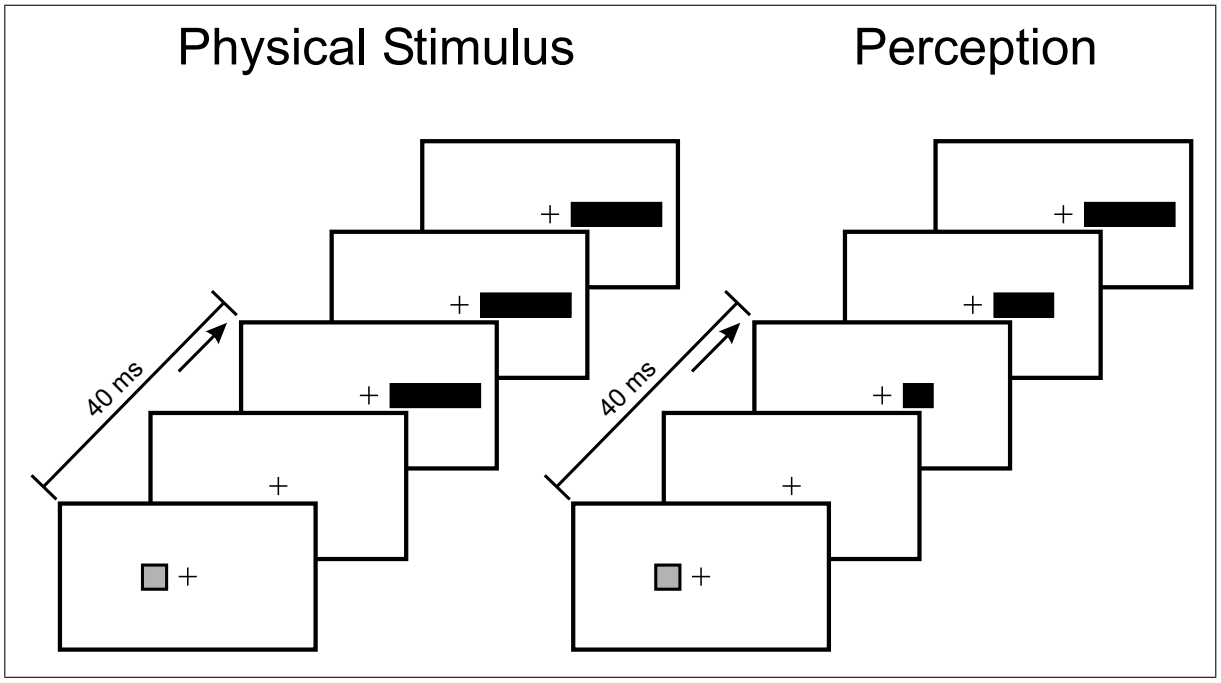


Abbildung 2.9: Erzeugung einer Bewegungsillusion durch fokale Aufmerksamkeit (nach: [Hikosaka et al., 1993a]): Einige 10 ms vor dem eigentlichen (Balken-)Reiz wird Fixpunkt für kurze Zeit ein Hinweisreiz eingeblendet. Beim Beobachter entsteht der Eindruck, daß der Balken aus dem Fixpunkt 'herauswächst', obwohl er in Wirklichkeit auf einmal eingeblendet wird. Der gleiche Effekt läßt sich auch ohne Hinweisreiz mit einem Kontrastgradienten erzeugen (Graukeil als Balken). Alle Effekte können durch einen der Wahrnehmung entgegengesetzten Zeitverlauf kompensiert werden; auf diesem Weg lassen sich die zeitlichen Veränderungen der Verarbeitung direkt messen.

Illusion wieder zum Verschwinden. Dies ist insbesondere wichtig, da auf diese Weise eine quantitative Analyse des Effekts möglich ist: Die wahrgenommene Zeitdifferenz ist vom Betrag gerade gleich der kompensierenden physikalischen Differenz.

Weitere Merkmalsdimensionen, in denen Latenzen auftreten, sind Intensitätskonstrast (stark vor schwach), Ortsfrequenz (grob vor fein, [Hughes et al., 1996]) und Orientierungskontrast v. Grünau et al. [1996b]. Die gleiche Bewegungsillusion läßt sich beispielsweise auch mit einem horizontalen Graukeil anstelle eines Hinweisreizes erzeugen: Die Seite mit dem stärksten Intensitätskonstrast wird zuerst wahrgenommen, der schwächere Kontrast verzögert. Darüber hinaus lassen sich beide Effekte gegeneinander ausspielen, d.h. sie können sich gegenseitig kompensieren [v. Grünau et al., 1996a,b]. Der Effekt kann außerdem durch mehrfache transiente Hinweisreize an mehreren Orten gleichzeitig und unabhängig hervorgerufen werden [Faubert und Grünau, 1995]. Über das reflektorische 'Einfangen' der Aufmerksamkeit durch präattentive Hinweisreize hinaus kann die zeitliche Wahrnehmung auch durch Suchaufgaben in ähnlicher Weise beeinflußt werden, wobei sich die einzelnen Beiträge teilweise anhand ihrer zeitlichen Charakteristik und ihres räumlichen Wirkungsbereichs unterscheiden lassen [v. Grünau et al., 1996b].

2.6 Modellierungs-Grundlagen

2.6.1 Das Marburger Modellneuron

In Kap. 1.1 wurde postuliert, daß sich alle wichtigen Prozesse der neuronalen Informationsverarbeitung auf der Ebene der Netzwerkdynamik verstehen lassen. Ausgehend von dieser Hypothese verwende ich eine Modellierung, die die mikroskopischen Prozesse an der Zellmembran nur summarisch betrachtet. Aktionspotentiale werden als pulsartige Ereignisse betrachtet, die am Zielneuron eine stereotype Reaktion auslösen (*Spike-Response*). Der Zeitpunkt, an dem ein Aktionspotential auftritt, wird allerdings mit hoher Zeitauflösung berücksichtigt, um Synchronisationseffekte nachbilden zu können.

Das verwendete Modell wurde von Eckhorn et al. [1990] vorgeschlagen und wird als *Marburger Modellneuron* bezeichnet. Die numerische Umsetzung in dieser Arbeit wurde bewußt mit beschränkter Rechengenauigkeit vorgenommen, um eine unproblematische Umsetzung auf die dedizierte Hardware zu ermöglichen (vgl. Kap. 3.1). Diese Variante wird im folgenden als *Accelerator-Neuron* bezeichnet und im folgenden Abschnitt erläutert.

Aus Abb. 2.10 ist die Struktur des Modells ersichtlich. Der dendritische Bereich integriert den von anderen Neuronen stammenden Spike-Input auf. Da die Spikes als zeitliche Delta-Funktionen modelliert werden, löst jeder Spike gerade die Impulsantwort der postsynaptischen Membran als PSP aus. Aufeinanderfolgende EPSPs werden linear überla-

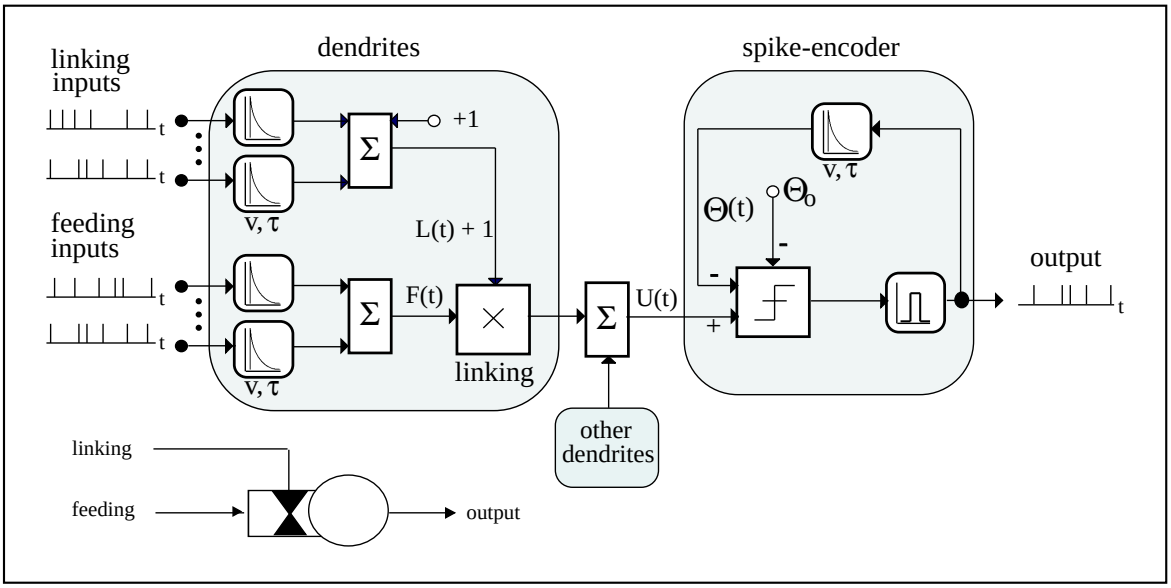


Abbildung 2.10: Struktur des Marburger Modellneurons. Zwei Dendritenzweige summieren unabhängig voneinander die PSPs auf, die von einlaufenden Spikes anderer Neuronen erzeugt werden. Am Soma werden die Teilpotentiale von Feeding- und Linking-Zweig getrennt aufaddiert und gemäß Gl. 2.2 zum Membranpotential verrechnet. Übersteigt das Membranpotential die Schwelle Θ, so wird für den betreffenden Zeitschritt ein Spike am Ausgang erzeugt. Nach jedem generierten Spike wird die Schwelle um einen konstanten Betrag V_Θ erhöht, um dann wieder auf ihren Ruhewert Θ_0 (Schwellenoffset) abzuklingen. Der so modellierte Refraktärmechanismus läßt sich formal auch als Selbstinhibition des Neurons auffassen. (Aus: [Eckhorn et al., 1990])

gert, so daß sich als dendritische Antwortfunktion die zeitliche Faltung von Eingangssignal und postsynaptischer Impulsantwort ergibt. Entsprechend dem in Abb. 2.2 angegebenen Ersatzschaltbild wird die Membran als Tiefpaß erster Ordnung behandelt und die entsprechende Impulsantwort als Faltungskern $\eta(t)$ bzw. EPSP verwendet:

$$\eta(t) = \begin{cases} 0 & : \quad t < 0 \\ w \cdot V \cdot e^{-\frac{t}{\tau}} & : \quad t \geq 0 \end{cases} \tag{2.1}$$

Von den Vorgängen am realen Neuron unterscheidet sich diese Modellierung insbesondere im verzögerungsfreien, sprunghaften Anstieg des postsynaptischen Membranpotentials beim Eintreffen des präsynaptischen Aktionspotentials (s. Abb. 2.3). Die Amplitude dieses Anstiegs wird durch die Verstärkung V und das *synaptische Gewicht* w der jeweiligen Neuronenverbindung bestimmt. Die Zeitkonstante τ, die das Abklingen der Membranspannung charakterisiert, ergibt sich aus den elektrischen Eigenschaften der Membran. Sie liegt in den Simulationen wie bei realen Neuronen im Millisekunden-Bereich.

Im Marburger Modellneuron finden zwei Typen von Dendriten Verwendung: Die *Feeding*-Synapsen, deren EPSPs additiv zum Membranpotential beitragen und die *Linking*-Synapsen, die eine modulatorische Funktion besitzen. Die entstehenden EPSPs werden an beiden Dendritentypen getrennt aufsummiert. Insbesondere können die Verstärkungen und Zeitkonstanten in beiden Zweigen verschieden sein. Diese werden mit den Indizes F und L für 'Feeding' bzw. 'Linking' bezeichnet.

Am Soma werden die beiden dendritischen Teilpotentiale zum Gesamt-Membranpotential verrechnet. Um der modulatorischen Funktion der Linking-Synapsen Rechnung zu tragen, werden diese multiplikativ mit einem Offset +1 hinzugefügt:

$$U(t) = F(t) \cdot (1 + L(t)) \tag{2.2}$$

mit

$$F(t) = \sum_{f=1}^{F} U_f(t) \qquad \text{und} \qquad L(t) = \sum_{l=1}^{L} U_l(t) \tag{2.3}$$

wobei F die Anzahl der Feeding- und L die Anzahl der Linking-Synapsen bedeutet.

Das so erhaltene Membranpotential wird mit der Feuerschwelle $\Theta(t)$ des Neurons verglichen. Diese hat im Ruhezustand den als *Schwellenoffset* bezeichneten Wert Θ_0. Ist das Membranpotential größer als die Schwelle, so wird am Ausgang ein Aktionspotential generiert; das Neuron feuert. Um nun die Refraktäreigenschaften natürlicher Neurone annähernd zu modellieren, kommt zum Schwellenoffset Θ_0 ein dynamischer Anteil hinzu, der nach jedem Spike um einen festen Wert V_Θ heraufgesetzt wird und dann exponentiell mit der Zeitkonstante τ_Θ abklingt (s. auch Abb. 2.4). Die gesamte dynamische Schwelle hat dann nach einem zur Zeit t_{Spike} ausgelösten Aktionspotential den folgenden Zeitverlauf:

$$\Theta(t) = \begin{cases} \Theta_0 & : \quad t < t_{Spike} \\ \Theta_0 + V_\Theta \cdot e^{-\frac{t}{\tau_\Theta}} & : \quad t \geq t_{Spike} \end{cases}$$

2.6.2 Das Acceleratorneuron

Die von FRANK ET AL. [1996] entwickelte Hardware zur Simulation des oben beschriebenen Neuronenmodells arbeitet alle zu simulierenden Neuronen in Folge ab, d.h. es handelt sich um eine *serielle Hardware* in FPGA-Technik.[1] Der Simulationsalgorithmus entspricht einem *synchronen Update* bei der herkömmlichen Simulation auf einem normalen Rechner, d.h. in einem festen Raster von Zeitschritten werden die Zustandsvariablen aller Neurone in Folge auf der Grundlage des vorhergehenden Zeitschritts berechnet. Deshalb wird dieses Verfahren als *Zeitschritt-Simulationsverfahren* bezeichnet. Abb. 2.11 zeigt schematisch den Ablauf von zwei aufeinanderfolgenden Zeitschritten: In der Erregungsphase werden alle von anderen Neuronen gesandten Impulse mit dem jeweiligen synaptischen Gewicht multipliziert und zu den jeweiligen Tiefpässen (Teilpotentialen) hinzuaddiert. Anschließend werden alle Tiefpässe mit ihrer jeweiligen Zeitkonstante abgeklungen. Treten Aktionspotentiale auf, so werden diese anschließend nach außen übertragen. Zusätzlich ist noch ein Lernen durch Veränderung der Synapsenstärken möglich.

Die wesentlichen Unterschiede des so modellierten Neurons zum Marburger Modellneuron bestehen in einer (teilweise stark) verminderten Rechengenauigkeit und einem leicht veränderten Abklingverhalten der postsynaptischen Potentiale. Letzteres kommt dadurch zustande, daß die Hardware nicht in jedem Zeitschritt alle Tiefpässe abklingt, sondern eine Liste der aktuell abzuklingenden Teilpotentiale verwaltet. Alle Teilpotentiale, deren Wert sich um weniger als eine Quantisierungsstufe (also den Gegenwert des geringstwertigen Bits im Speicher) ändert, werden als Null angesehen und aus der Abklingliste gestrichen. Dies bringt einerseits eine erhebliche Ersparnis an Rechenzeit, führt aber andererseits dazu, daß jeder Tiefpaß immer um mindestens eine Quantisierungsstufe erniedrigt werden

[1] FPGA steht für **F**ield **P**rogrammable **G**ate **A**rray. Diese Hardware-Technologie verwendet frei programmierbare Logik-Schaltungen und erlaubt dadurch ein schnelles und flexibles Erstellen neuer Hardware-Designs. Insbesondere entfällt der aufwendige Layout- und Maskenprozeß, der bei der Herstellung anwendungsspezifischer integrierter Schaltungen (ASIC) notwendig ist. Die größere Flexibilität von FPGA-Designs wird allerdings mit generellen Geschwindigkeitseinbußen erkauft.

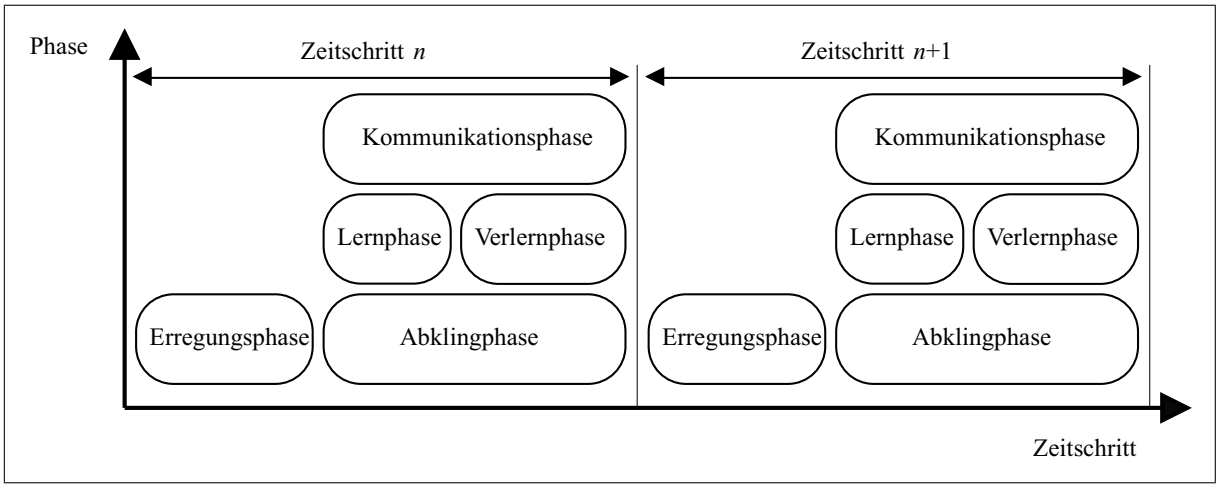


Abbildung 2.11: Ablauf zweier aufeinanderfolgender Zeitschritte bei der Simulation in Hardware, Details im Text. (Nach: FRANK ET AL. [1996])

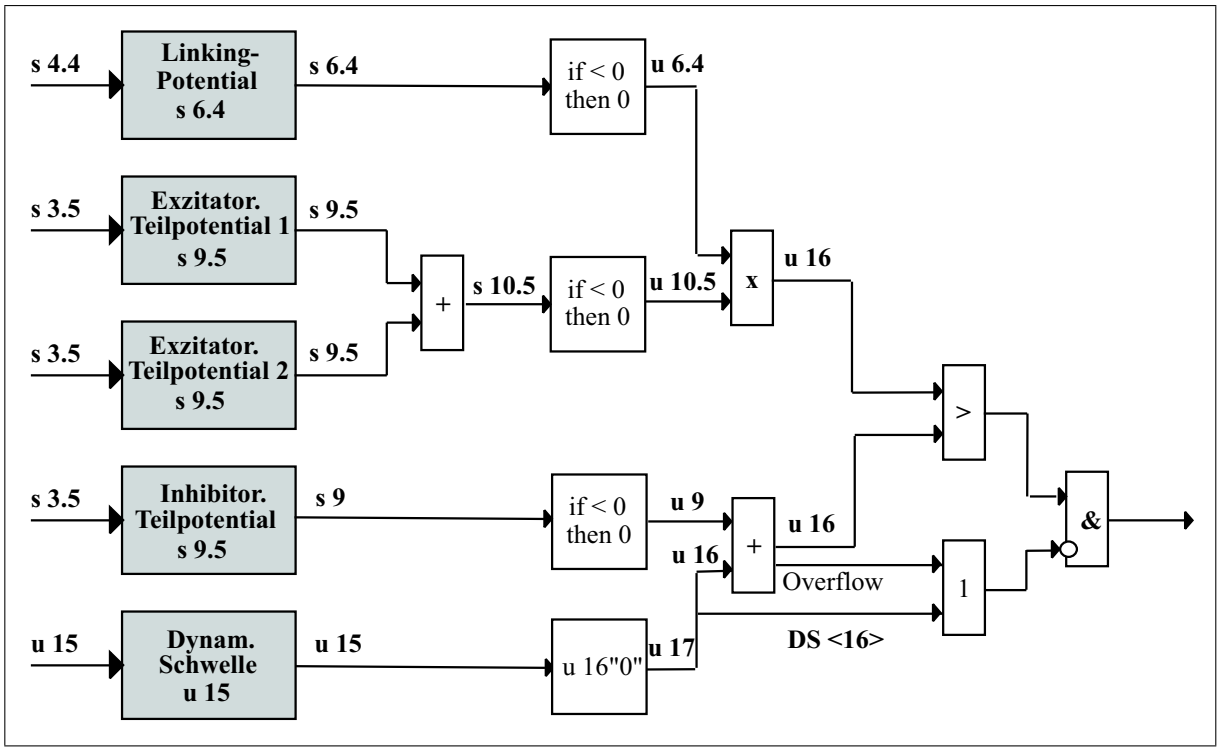


Abbildung 2.12: Aufbau des Acceleratorneurons (nach: FRANK ET AL. [1996]). Die Teilpotentiale entsprechen den im vorigen Abschnitt besprochenen Größen des Marburger Modellneurons; die Rechengenauigkeit der Teilpotentiale und Zwischengrößen ist in Bits vor und nach dem Dualkomma angegeben. Mit **u** (wie *unsigned*) bezeichnete Größen tragen keine Vorzeicheninformation; vorzeichenbehaftete Größen sind mit **s** (wie *signed*) gekennzeichnet. Beispiel: *s 3.5* = vorzeichenbehaftete Größe mit 8+1 Bit Länge. Von den 8 Bits, die den Betrag festlegen, befinden sich 3 vor und 5 nach dem Dualkomma; das höchste Bit entspricht also 2^2, das niedrigste 2^{-5}.

muß, da er sonst aus der Abklingliste gestrichen wird und das Potential nicht auf Null zurückgehen kann. Der letzte, flache Teil der Exponentialfunktion aus Gl. 2.1 wird also zu einem linearen Abfall 'verbogen'. In Abb. 2.12 sind die Rechengenauigkeiten für die einzelnen Teilpotentiale zusammengefaßt. Daraus geht auch hervor, daß nur eine begrenzte Anzahl von Zeitkonstanten für die verschiedenen Synapsentypen zur Verfügung steht. Dies erweist sich als deutliche Einschränkung, wenn Teile des magno- und des parvozellulären Systems gemeinsam betrieben werden sollen, da sich ihre zeitlichen Eigenschaften deutlich unterscheiden.

2.6.3 Das Neuronenmodell von McCulloch und Pitts

Da in Kap. 6 eine stationäre Näherung behandelt wird, die sich auf das Neuronenmodell von MCCULLOCH und PITTS [1943] abbilden läßt, stelle ich dieses ebenfalls kurz vor. Abb. 2.13 zeigt den Aufbau dieses Modells. Ebenso wie ein natürliches Neuron besteht es aus einem dendritischen Eingangsbereich, einem Soma, an dem die Eingangssignale integriert werden, und einem binären Ausgang, der dem Axon entspricht. Alle Eingangssignale x_i werden im Soma zum Membranpotential U addiert. Dieser Wert wird mit einer festen Schwelle Θ verglichen, die der Feuerschwelle des Neurons entspricht. Ist die Summe der

Eingangswerte größer als die Schwelle, so wird der Ausgang Y auf 1 gesetzt, andernfalls auf 0:

$$U = \sum_{i=1}^{n} u_i \tag{2.4}$$

$$Y = \begin{cases} 0 & : \quad U < \Theta \\ 1 & : \quad U \geq \Theta \end{cases} \tag{2.5}$$

Diese stark vereinfachte Modellierung berücksichtigt weder eine eventuelle nichtlineare Integration der Eingangssignale noch eine innere Dynamik des Neurons. Modelliert wird lediglich die qualitative Eigenschaft der Über- oder Unterschwelligkeit des Neurons. Trotzdem stellt diese bereits ein wesentliches Element in der Arbeitsweise realer Neurone dar – auch reale Neurone produzieren keine Ausgangsaktivität, solange sie nicht über die Schwelle hinaus depolarisiert werden.

Für die Verwendung dieses Neuronenmodells spricht neben der einfachen numerischen Behandlung als wichtigster Grund die Möglichkeit, analytische Rechnungen durchzuführen. Wie sich in Kap. 4.1 noch zeigen wird, ist dies mit komplexeren Neuronenmodellen nur noch in hochgradig idealisierten Spezialfällen machbar.

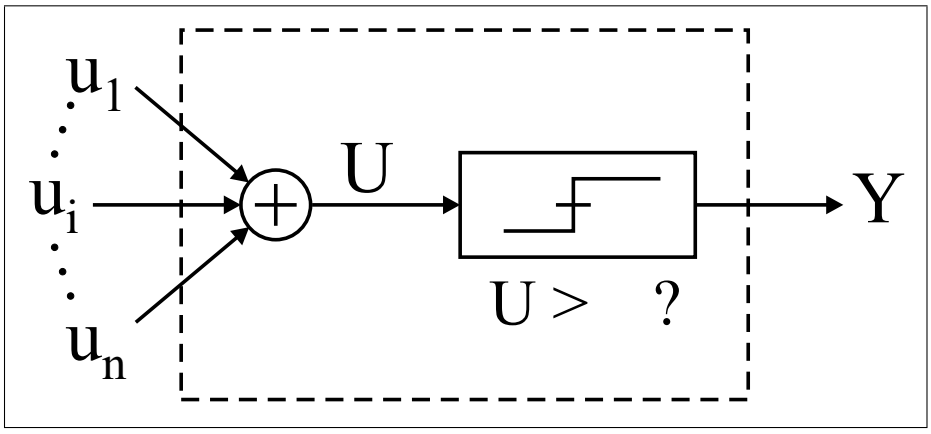


Abbildung 2.13: Das Neuronenmodell von McCulloch und Pitts [1943]. Alle Eingangssignale werden aufsummiert durch Vergleich mit einer festen Schwelle Θ zu einem binären Ausgangswert verrechnet. Die gestrichelte Linie deutet die Zellmembran des Neurons an.

3 Struktur und Eigenschaften des Aufmerksamkeitssystems

3.1 Überblick

Das Gesamtsystem (s. Abb. 3.1) gliedert sich in vier Teile: Vorverarbeitung, Kontur-Form-System, transientes System und Aufmerksamkeitssteuerung. Die Vorverarbeitung codiert die lokalen Intensitätsunterschiede im Bild in Spikefolgen, die jeweils als Eingangssignal für das Kontur-Form- und das Transientensystem dienen. Während das in Abschnitt 3.3 beschriebene Kontur-Form-System daraus (quasi-)stationäre räumliche Merkmale extrahiert und diese zur Segmentierung verwendet, erkennt der transiente Zweig bewegte Bildteile und codiert diese mit Bewegungsrichtung und -geschwindigkeit in einen Satz retinotoper Merkmalskarten. Die so gewonnene, im räumlichen Sinn spärliche Information liefert mögliche Positionen bewegter Objekte (zusammen mit der geschätzten Bewegungsrichtung). Auf der Grundlage dieser *Vorauswahl* wählt die Aufmerksamkeitssteuerung ein eindeutiges Blickziel aus. Zusätzlich wird die Information über die Bewegungsrichtung für die weitere Verfolgung eines Objektkandidaten herangezogen.

Wie in Kap. 5 gezeigt wird, lassen sich mit dieser Systemarchitektur Objekte bzw. Objektkandidaten praktisch ohne Vorwissen erfassen und verfolgen, so daß trotz eines nichtstationären Eingangssignals eine Segmentierung im Kontur-Form-System möglich ist. Die Einzelheiten hierzu werden in Kap. 4 erläutert.

3.2 Die Vorverarbeitung: Modellierung der Retina

3.2.1 Räumliche Eigenschaften

Wie in Kap. 2.3.2 vorgestellt, codieren bereits die Ganglienzellen am Ausgang der Retina nicht mehr einfache Intensitäts- bzw. Farbsignale, sondern antworten primär auf lokale Kontraste im retinalen Bild. Im natürlichen System wird diese Kontrastempfindlichkeit von einer weiteren Zellschicht, den *Bipolarzellen*, erzeugt. Diese haben als Ausgangssignal ein graduiertes Potential und speisen damit die Ganglienzellen. Für die Modellierung bringt eine explizite Berücksichtigung der Bipolarzellen allerdings keinen Vorteil; im Gegenteil läßt sich (bei praktisch gleichem Ergebnis) erheblich Rechenzeit einsparen, wenn die Rezeptoren direkt mit einer geeigneten räumlichen Struktur auf die Ganglienzellen aufgeschaltet werden. Neben den räumlichen werden vom natürlichen System wie vom Modell noch zeitliche Eigenschaften des Reizes in die Folge der Aktionspotentiale, d.h. die Intervalle zwischen ihnen eincodiert.

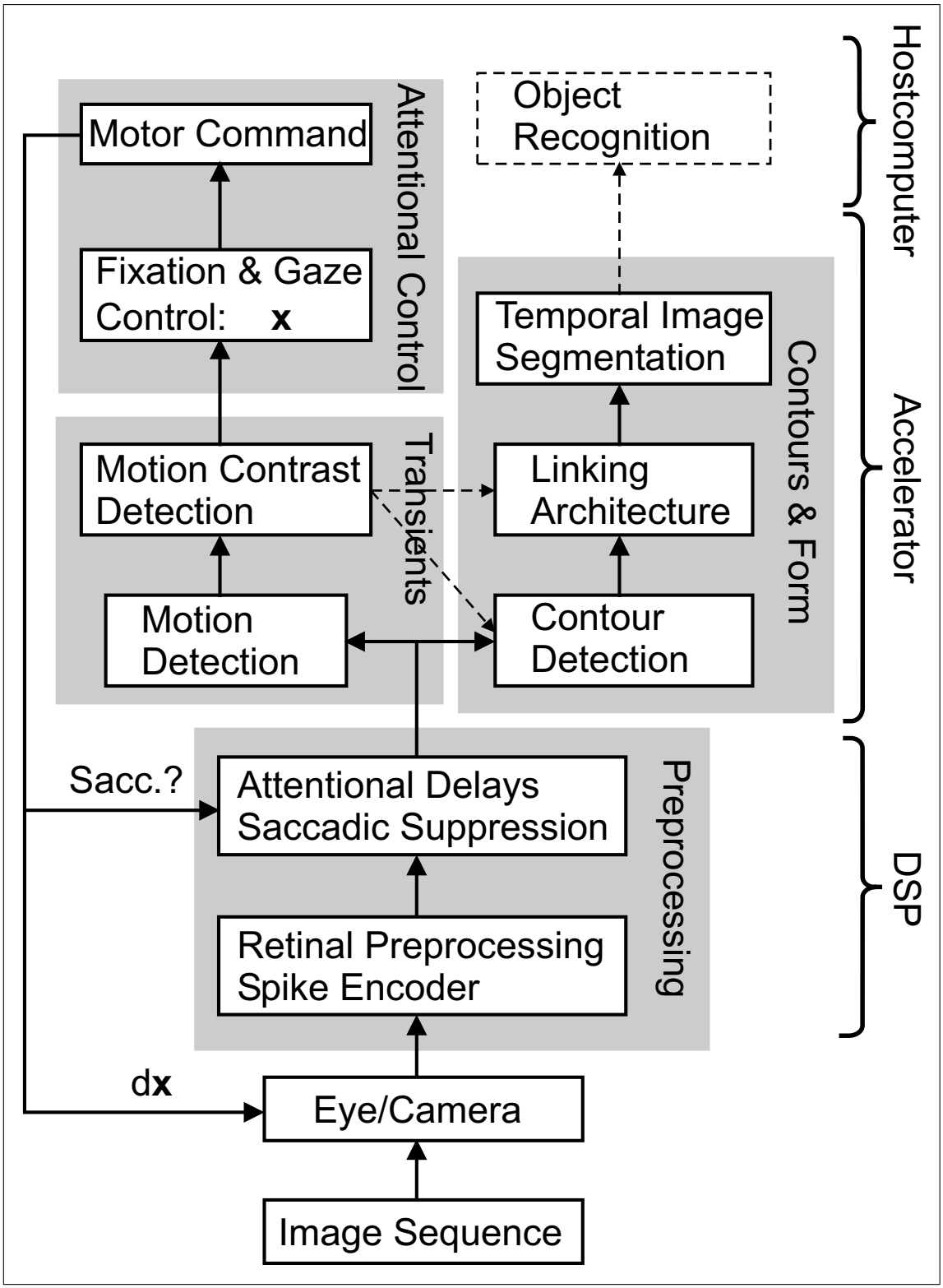


Abbildung 3.1: Überblick über die Struktur des gesamten Systems. Die von der retinalen Vorverarbeitung generierte Spikefolge wird dem Transienten- und dem Kontur-Formsystem zugeführt. Die vom transienten Zweig extrahierte Bewegungsinformation wird vom Aufmerksamkeitsmodul ausgewertet und zur Kamerasteuerung verwendet, so daß eine geschlossene Regelschleife entsteht. Am rechten Rand ist die vorgesehene Rechnerplattform der Module im Echtzeit-Betrieb angegeben.

Die räumliche Abhängigkeit der Zellantwort vom Intensitätsmuster ist durch eine *Center-Surround-Organisation* gekennzeichnet, wobei man ON-Center und OFF-Center Zellen unterscheidet. Eine ON-Center-Zelle ist dadurch charakterisiert, daß Lichteinfall im Zentrum ihres rezeptiven Feldes die Zelle aktiviert, Lichteinfall in einem ringförmigen Bereich um das Zentrum sie dagegen hemmt. Beide Bereiche zusammen füllen das rezeptive Feld aus, wobei die Gewichtung gerade so angelegt ist, daß bei der Präsentation eines großflächigen Reizes, der das gesamte RF bedeckt, keine dauernde Aktivierung gegenüber dem Ruhezustand auftritt (zum Geschehen beim Einschalten des Reizes s. folgenden Abschnitt). Diese Symmetrie zwischen ON- und OFF-Bereich des rezeptiven Feldes gilt in umgekehrter Form für die Off-Center/On-Surround-Zellen; diese antworten somit auf Dunkelheit im Zentrum und Helligkeit in der Peripherie.

Der Aktivierungsbeitrag der verschiedenen RF-Bereiche läßt sich empirisch in guter Näherung durch eine sogenannte Mexikanerhut-Funktion mit umgekehrtem Vorzeichen beschreiben. Diese ergibt sich im eindimensionalen Fall als zweite Ableitung einer Gaußfunktion mit Breite σ. Im zweidimensionalen Fall tritt an die Stelle der zweiten Ableitung der Laplace-Operator. Mit der zweidimensionalen, isotropen Gaußfunktion der Breite σ

$$G_\sigma(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{3.1}$$

ergibt sich also für die Mexikanerhut-Funktion

$$M(x,y) = \nabla^2 G_\sigma(x,y) = \frac{2\sigma^2 - (x^2+y^2)}{2\sigma^6\pi} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.2}$$

Abb. 3.2 zeigt das Profil dieser Funktion am Beispiel einer ON-Center-Zelle. Das Integral über den exzitatorischen (positiven) Innenbereich entspricht gerade dem über das inhibitorische Umfeld (die 'Krempe' des Mexikanerhuts), so daß die oben beschriebene Symmetrie entsteht. Nach außen hin fällt die Funktion auf Null ab; dort endet das rezeptive Feld der Zelle.

Abb. 3.3 illustriert, was geschieht, wenn sich im RF einer ON-Zelle eine Kante, d.h. ein Hell-Dunkel-Übergang befindet. Die exzitatorische Reizung des Zentrums wird – anders als beim homogenen Reiz – nicht vollständig durch eine gleich starke Inhibiton aus dem Umfeld ausgeglichen, so daß die Zelle aktiviert wird und überschwellig werden kann.

Setzt man an Stelle der ON- eine OFF-Zelle an die gleiche Position, so überwiegt die aus dem Zentrum stammende Inhibition die Exzitation aus dem Umfeld; die OFF-Zelle wird also gehemmt. Kehrt man die Kontrastrichtung der Kante um, so vertauschen ON- und OFF-Zellen ihre Rollen.

Die gemeinsame Aktivität von ON- und OFF-Zellen codiert also das Vorhandensein und die Kontrastpolarität von Hell-Dunkel-Übergängen in ihrem jeweiligen RF. Sie lassen allerdings noch keinen eindeutigen Rückschluß auf die Orientierung einer eventuellen Kante zu, die den Übergang verursacht. Im Abschnitt 3.3.1 wird erläutert, wie durch geeignete Überlagerung von ON- und OFF-Zellen Kanten einer bestimmten Orientierung und Kontrastpolarität detektiert werden können.

Die Orte der Rezeptorzellen in der Retina liegen in guter Näherung auf den Schnittpunkten eines hexagonalen Gitters. In der *Fovea Centralis*, dem zentralen Bereich des

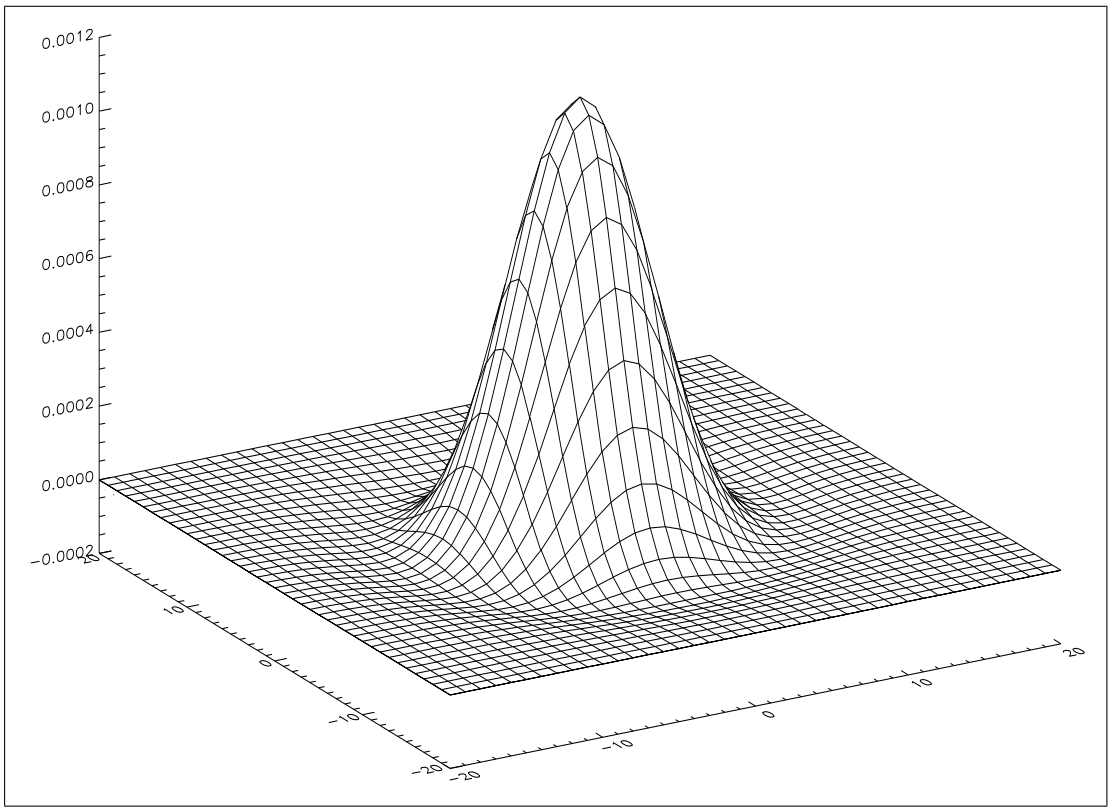


Abbildung 3.2: Die nach ihrer Form benannte 'Mexikanerhut'-Funktion, die als näherungsweise Beschreibung für das räumliche Profil der rezeptiven Felder von Ganglienzellen dient. Innen- und Außenbereich sind im Integral gerade gleich groß, so daß bei homogener (flächiger) Reizung die Zelle nicht aktiviert wird.

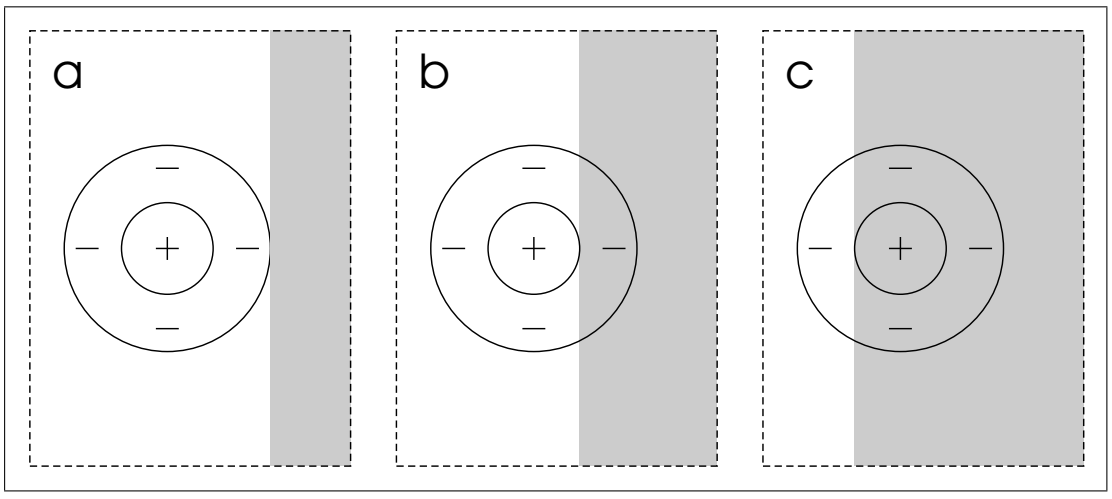


Abbildung 3.3: Wirkung eines Hell-Dunkel-Übergangs im RF einer ON-Center-Ganglienzelle. Die exzitatorische Reizung des Zentrums wird – anders als beim homogenen Reiz – nicht vollständig durch eine entsprechende Inhibiton aus dem Umfeld ausgeglichen, so daß die Zelle depolarisiert wird und überschwellig werden kann.

Gesichtsfeldes, ist die Gitterkonstante am kleinsten und somit die räumliche Auflösung die höchste im Sehfeld. Außerhalb der Fovea nimmt die Rezeptordichte etwa logarithmisch mit dem Abstand vom Zentrum ab, wobei die rezeptiven Felder in gleichem Maß größer werden. Mit der Abnahme der räumlichen Auflösung geht auch eine Veränderung der zeitlichen Eigenschaften einher; diese werden im folgenden Abschnitt erläutert.

3.2.2 Zeitliche Eigenschaften

3.2.2.1 Zeitliche Eigenschaften der Modell-X-Zellen

Obwohl das Prinzip der lokalen Kontrastempfindlichkeit in Form einer räumlichen Center-Surround-Organisation allen Ganglienzellen gemeinsam ist, unterscheiden sie sich in ihren zeitlichen Eigenschaften. Wie bereits in Abschnitt 2.3.1 dargestellt, unterscheidet man langsame X- und schnelle Y-Neurone. Erstere sind als parvozelluläre Neurone für die Konturverarbeitung und -erkennung geeignet, während die zweite Gruppe dem magnozellulären Pfad, also der Bewegungs- und Transientenwahrnehmung zugerechnet wird.

Die X-Zellen zeigen neben der oben beschriebenen räumlichen Symmetrie zwischen Zentrum und Umfeld ihrer RFs eine zeitliche Asymmetrie: Der zentrale RF-Bereich antwortet (bei gleicher stationärer Gewichtung) generell schneller als das Umfeld, so daß beim Einschalten eines homogenen Reizes die Exzitation aus dem RF-Zentrum kurzzeitig überwiegt und auch ohne lokalen Hell-Dunkel-Übergang eine transiente Antwort der Zelle entsteht. Dieses Verhalten läßt sich gut durch Verwendung unterschiedlicher Zeitkonstanten an den Feeding-Eingängen von Zentrum und Peripherie modellieren. Abb. 3.4 verdeutlicht das Zusammenspiel der beiden Signalanteile. Bei länger dauernder Präsentation eines homogenen Flächenreizes nehmen die Potentiale an beiden Tiefpässen den gleichen Sättigungswert an, so daß die Zelle effektiv nicht mehr aktiviert wird.

Dieser Effekt macht sich auch beim Einschalten einer ausreichend kontraststarken

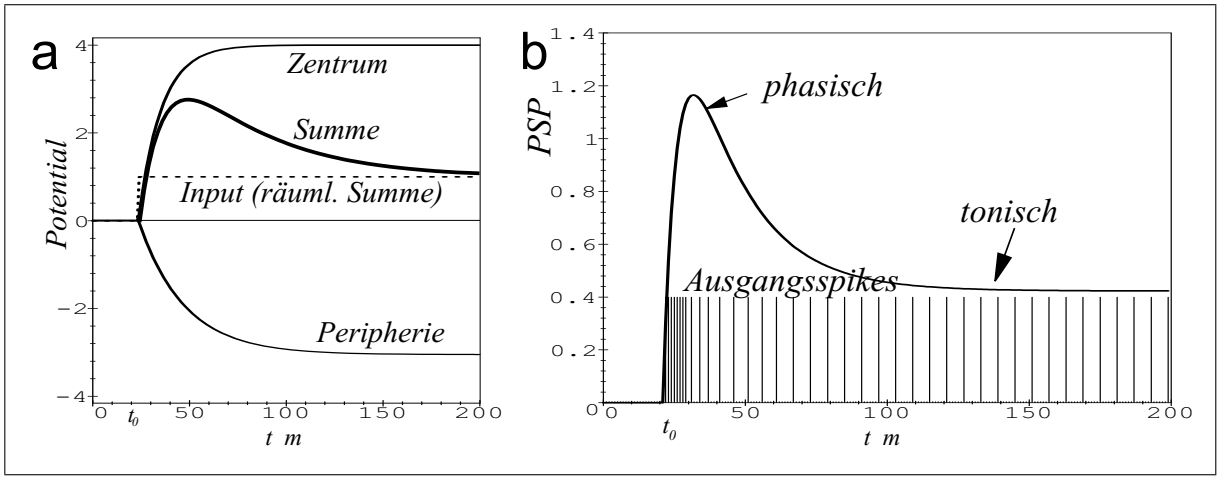


Abbildung 3.4: Zeitliche Antworteigenschaften der X-Zellen. Zentrum und Umfeld werden mit gleichem Gewicht, aber unterschiedlichen Zeitkonstanten im Leckintegrator addiert, so daß unmittelbar nach dem Einschalten die zentrale Exzitation stärker wirksam wird als die Inhibition aus dem Umfeld des RF. Die vorübergehende Verstärkung wird auch als *phasische*, der dauerhafte Anteil als *tonische* Antwort bezeichnet.

Hell-Dunkel-Kante bemerkbar. Hier ist zwar die stationäre Aktivierung verschieden von Null – aber durch die schnellere Antwort des Zentrums tritt auch hier unmittelbar nach dem Einschalten eine kurzfristig höhere Aktivität der Zelle auf (*phasische Antwort*), die bei längerer Reizdarbeitung in die niedrigere stationäre Feuerrate (*tonische Antwort*) übergeht. Dieses Antwortverhalten führt gleichzeitig zu einer gewissen Bewegungsempfindlichkeit der X-Zellen; die Einzelheiten dazu werden in s. Kap. 3.4 besprochen.

Im verwendeten Simulationsprogramm sind prinzipiell für Zentrum und Umfeld verschiedene Zeitkonstanten vorgesehen; einige Simulationen, die lediglich auf der Detektion von Intensitätskontrasten basieren, wurden jedoch mit gleichen Zeitkonstanten für Zentrum und Umfeld durchgeführt.

3.2.2.2 Zeitliche Eigenschaften der Y-Zellen

Im natürlichen Sehsystem unterscheiden sich die Y-Zellen im wesentlichen durch zwei Merkmale von den X-Zellen:

1. Sie haben größere rezeptive Felder als X-Zellen, deren Durchmesser mit wachsender Exzentrizität zunimmt. Als Folge davon ist das retinale Auflösungsvermögen in der Peripherie deutlich schlechter als im Zentrum.

2. Sie reagieren in erster Linie auf transiente Ereignisse bzw. Bewegung. Stationäre Reize rufen dagegen nur eine schwache oder gar keine Antwort hervor.

Zwar zeigen auch X-Zellen eine vorübergehend verstärkte Antwort beim Einschalten eines Reizes; dies spielt jedoch eher die Rolle eines zusätzlichen Signalanteils. Bei den Y-Zellen macht im Gegensatz dazu die transiente Zellantwort den entscheidenden Teil aus; dementsprechend baut die gesamte Bewegungsverarbeitung des Modells auf dem Output der Y-Zellen auf. Im Modell läßt sich ein transientes Antwortverhalten – ähnlich wie bei den X-Zellen – durch eine Überlagerung zweier Tiefpässe mit verschiedenen Zeitkonstanten und entgegengesetztem Vorzeichen erreichen. Allerdings wird man zur Vermeidung einer stationären Antwort *dasselbe* Eingangssignal auf beide Tiefpässe geben; so ist sichergestellt, daß unabhängig von der räumlichen Struktur des Reizes die Aktivierung im gesättigten Zustand verschwindet. Als Eingangssignal für einen solchen *Transientendetektor* eignet sich sowohl ein direkt durch Faltung mit einem RF-Profil errechneter Wert als auch der Spike-Output einer X-Zelle. Beide Varianten kamen in der Modellierung zum Einsatz; die Einzelheiten sind in Kap. 3.4 angegeben. Abb. 3.5 veranschaulicht das Zustandekommen der transienten Antwort.

3.2.3 Modellierung und technische Umsetzung

3.2.3.1 Allgemeines

Für eine technische Umsetzung des Modells im Computer muß das Eingangssignal vollständig diskretisiert werden. Dies geschieht durch Abtastung in zwei Raumrichtungen sowie auf der Zeitachse. Schließlich ist auch die Amplitudendarstellung quantisiert. Um

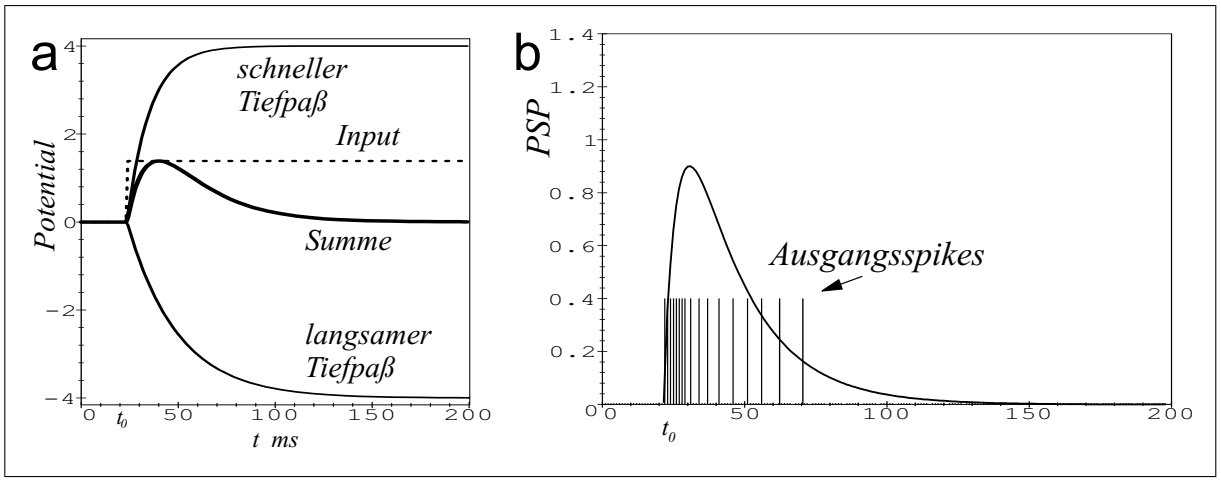


Abbildung 3.5: Zeitliche Antworteigenschaften der Y-Zellen. Dasselbe Eingangssignal wird mit unterschiedlichen Zeitkonstanten tiefpaßgefiltert und der langsamere der beiden Signalanteile vom schnelleren subtrahiert. Jede ausreichend schnelle zeitliche Änderung im Eingangssignal führt so zu einer Antwort, die wieder erlischt, wenn auch der langsame Tiefpaß die Sättigung erreicht hat.

dabei Fehler durch eine Unterabtastung zu vermeiden, muß ggf. vorher eine entsprechende Tiefpaßfilterung durchgeführt werden. Alle Koordinatenachsen erhalten sinnvollerweise ganzzahlige Maßeinheiten: Die Ortskoordinaten werden in Bildpunkten (Pixel) gemessen, als Einheit der Zeitmessung dient ein Simulationsschritt (1 bin). Die Grauwerte der einzelnen Bildpunkte werden als natürliche Zahlen zwischen 0 und 255 dargestellt, wobei 0 schwarz und 255 weiß codiert. Dies entspricht einer Amplitudenquantisierung mit einer Auflösung von 8 bit.

3.2.3.2 Zeitliche Abtastung und Interpolation

Da die Bildaufnahme mit einer handelsüblichen CCD-Videokamera (schwarzweiß) erfolgt, liegen die Eingangsbilder als Folge von Zeilenhalbbildern im Abstand von 20 ms vor. Davon wird nur jedes zweite ausgewertet, da ansonsten der räumliche Versatz zwischen zwei aufeinanderfolgenden Halbbildern berücksichtigt werden müßte, was erheblichen rechnerischen Mehraufwand bedeutet. Der Abstand zweier Eingangsbilder (auch als *Frames* bezeichnet) beträgt also $\Delta t_{Frame} = 40\,ms$, was einer Abtastfrequenz von $f_{Frame} = 25\,Hz$ entspricht. Um die Simulation mit der Eingangsbildfolge zu synchronisieren, wird Δt_{Frame} gleich 32 bin gesetzt:

$$1\,bin = \frac{1}{32}\Delta t_{Frame} = 1.25\,ms \tag{3.3}$$

Wie in Kap. 2.6 erläutert, ist eine Zeitauflösung in dieser Größenordnung für die Simulation neuronaler Vorgänge erforderlich. Im Vergleich zur Bildfolge wird das Eingangssignal also mehrfach überabgetastet. Dies stellt zwar im Hinblick auf die Signalverarbeitung kein Problem dar, wirft jedoch die Frage nach einer geeigneten Rekonstruktion der unbekannten Grauwertbilder zwischen den von der Kamera gelieferten Frames auf. Die einfachste Lösung, bis zum Eintreffen des nächsten Kamerabildes die 'alten' Grauwerte weiter zu verwenden, führt dazu, daß sich alle 40 ms das Potential am Eingang aller

Rezeptorzellen sprunghaft ändert, dazwischen jedoch konstant bleibt. Ein solcher Signalverlauf stellt für ein System, das die Zeit selbst als Codierungsdimension verwendet und durch seine interne Dynamik regelmäßige Oszillationen und somit einen Takt generieren soll, eine starke Beeinflussung dar. Diese Vermutung, daß ein derartiger externer Takt die Netzwerkdynamik stark beeinflussen könnte, wurde in Voruntersuchungen, bestätigt [Stöcker, 1994].

Die theoretisch beste Möglichkeit, im Rahmen des Abtasttheorems das ursprüngliche Signal zu rekonstruieren, bietet eine Interpolation mit einem $\sin(t)/t$-Kern [Lüke, 1975]. Dieses Verfahren benötigt zur optimalen Rekonstruktion aber den gesamten Zeitverlauf des Eingangssignals. Dies stellt bei einer Interpolation, die im nachhinein vorgenommen wird, kein Problem dar. Für ein echtzeitfähiges System, dessen Betriebsdauer zudem beliebig lang sein kann, muß man sich dagegen mit einer teilweisen Rekonstruktion zufrieden geben. Die Güte der Interpolation steigt dabei mit der Anzahl der ausgewerteten Stützstellen. Da diese immer symmetrisch um den zu rekonstruierenden Zeitschritt liegen (d.h. jeweils die Hälfte in der 'Vergangenheit' und 'Zukunft' des Meßpunkts), muß bei einer größeren Zahl von ausgewerteten Stützstellen auch länger gewartet werden, bis mit der Berechnung begonnen werden kann. Die dadurch erzwungene Verzögerung bei der Erzeugung des Eingangssignals stellt für ein echtzeitfähiges System ein echtes Problem dar, das nur durch einen Kompromiß zwischen Rekonstruktionsgüte und Zeitversatz gelöst werden kann.

Als Ausweg bietet sich hier eine einfache, lineare Interpolation zwischen den von der Kamera bereitgestellten Stützstellen an. Dabei muß nur ein Kamerabild 'aus der Zukunft' berücksichtigt werden, d.h. das Signal wird mit einer Verzögerung von $\Delta t_{Frame} = 40\,ms$ an das neuronale System weitergereicht. Ein weiterer Vorteil ist die einfache rechnerische Umsetzung, die der geforderten Echtzeitfähigkeit stark entgegenkommt. Zudem zeigte sich in den o.a. Voruntersuchungen, daß die zusätzlichen hochfrequenten Signalanteile, die durch die lineare Interpolation zwangsläufig eingeführt werden, keine gravierenden Auswirkungen haben. Letzteres ist in erster Linie auf die zeitlichen Tiefpaßeigenschaften am Eingang der verwendeten Modellneurone zurückzuführen. Zu wahrnehmbaren Unterschieden in der Netzwerkdynamik führt erst ein globaler Takt, der für alle Neurone von außen aufgeprägt wird.

Selbstverständlich ist mit der zeitlichen Abtastung eine obere Grenzfrequenz von $f_{Frame}/2 = 12.5\,Hz$ verbunden. Diese ist für die hier geforderten Leistungen völlig ausreichend; eine höhere Frequenz ließe sich durch den Einsatz einer schnelleren Kamera ohne Probleme realisieren.

3.2.3.3 Räumliche Abtastung

Die Anordnung der natürlichen Rezeptorzellen in der Retina läßt sich in guter Näherung durch ein hexagonales 'Bienenwaben'-Gitter beschreiben, das die höchste Packungsdichte in zwei Dimensionen ermöglicht. Die einzelnen Rezeptoren haben ein annähernd gaußförmiges räumliches Empfindlichkeitsprofil, das sich mit denen der nächsten Nachbarn etwa auf halber Höhe des Maximums schneidet. Die exakte Nachbildung eines solchen Gitters ist aufwendig, da die Punkte sich nicht auf ganzzahlige kartesische Koordinaten

abbilden lassen. Eine Anhäherung läßt sich jedoch durch eine Anordnung realisieren, die zwar eine hexagonale Grundstruktur besitzt, deren Punkte aber mit denen des kartesischen Gitters zusammenfallen [Hartmann, 1982]. Abb. 3.6 illustriert diese Art der Anordnung, die im folgenden als *pseudohexagonales Gitter* bezeichnet wird. Jedes pseudohexagonale Gitter ist durch seine *Abtastweite*, d.h. den kleinsten Rezeptorabstand d eindeutig gekennzeichnet. In der Abbildung sind die Neuronenpositionen jeweils für eine Rezeptorschicht mit $d = 2$ und $d = 4$ dargestellt. Die Winkel der Hauptachsen des Gitters gegen die Horizontale betragen 63.4° bzw. 116.6°. Der einfacheren Notation halber bezeichne ich diese Achsen im folgenden mit ganzzahligen Vielfachen von 30°; gemeint sind jeweils die tatsächlichen Orienterungen im pseudohexagonalen Gitter.

Formal läßt sich das pseudohexagonale Raster mit der Abtastweite d als binärwertige Funktion der Pixelkoordinaten beschreiben; Bildpunkte, die mit einer Rezeptorposition zusammenfallen, erhalten den Wert 1, alle anderen 0:

$$R(i,j) = \begin{cases} 1 & : \quad (i \; mod \; d = 0 \; \wedge \; j \; mod \; 2d = 0) \; \vee \; (i \; mod \; d = \dfrac{d}{2} \; \wedge \; j \; mod \; 2d = d) \\ 0 & : \quad sonst \end{cases} \tag{3.4}$$

wobei *mod* den Rest der Ganzzahldivision bezeichnet.

Diese Anordnung läuft also auf eine räumliche Unterabtastung des Eingangsbildes hinaus. Um das Abtasttheorem nicht zu verletzen, muß das Eingangsbild vor der Faltung tiefpaßgefiltert, d.h. geglättet werden. Dies geschieht – in Übereinstimmung mit dem o.g. räumlichen Empfindlichkeitsprofil der Rezeptoren – durch Falten mit einer normierten Filtermaske mit gaußförmiger Gewichtung der einzelnen Pixel. Die Normierung stellt sicher, daß die Gesamthelligkeit des bearbeiteten Bildbereichs nicht verändert wird. Ferner ist darauf zu achten, daß alle Pixel gleich stark zur Entstehung der Rezeptorpotentiale beitragen. Eine ausführliche Diskussion dieser Problematik findet sich in [Spengler, 1996].

Ist die Standardabweichung der gaußförmigen Gewichtung in der Filtermaske bekannt, so ist diese eindeutig definiert; die Einträge werden durch Abtasten der zweidimensionalen Gaußfunktion bestimmt:

$$Z_\sigma(k,l) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}\left(\frac{k^2+l^2}{2\sigma^2}\right)} \tag{3.5}$$

Die für verschiedene Abtastweiten d verwendeten Werte für σ sind in Tab. 3.1 aufgelistet. M bezeichnet dabei die (immer ungerade) Anzahl der Spalten bzw. Zeilen der jeweiligen Gaußmaske. Der übersichtlicheren Notation halber definieren wir $m = (M-1)/2$; damit wird der Abstand in Pixeln vom zentralen Element der Gaußmaske zu ihrem Rand angegeben. Die Rezeptorpotentiale $P(i,j)$ werden durch Faltung der Eingangsgrauwerte $I(i,j)$ mit der Gaußmaske $Z_\sigma(k,l)$ ermittelt:

$$P(i,j) = \sum_{k=-m}^{m} \sum_{l=-m}^{m} I(i,j) Z_\sigma(i-k, j-l) \text{ für alle } R(i,j) = 1 \tag{3.6}$$

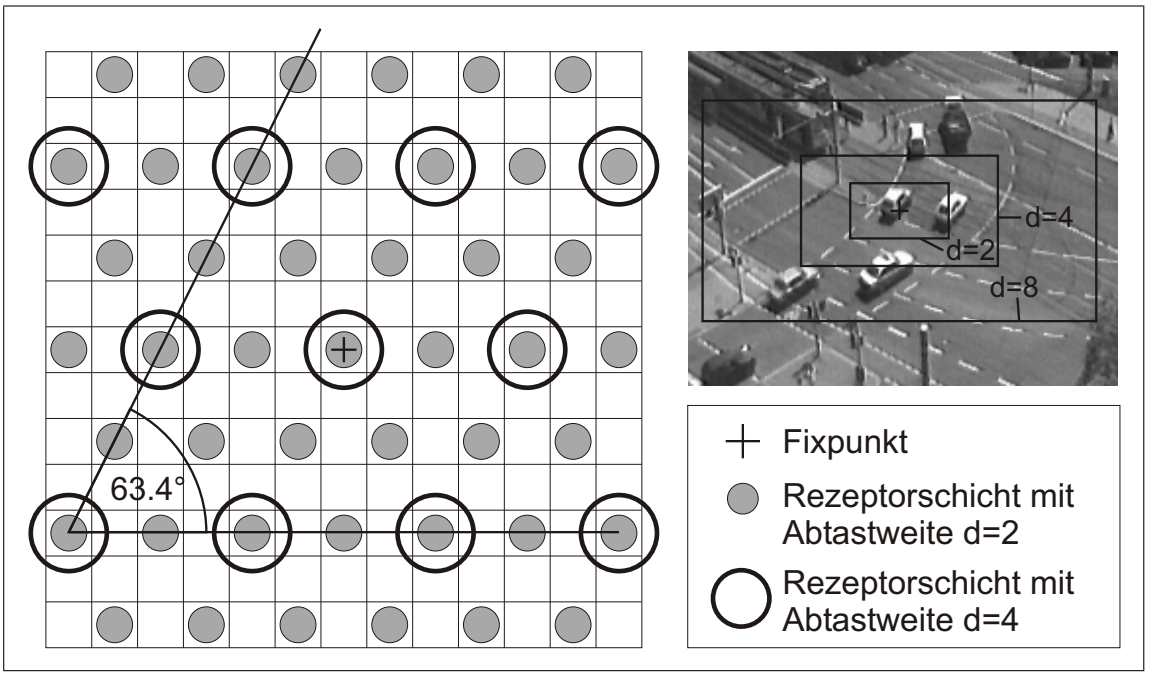


Abbildung 3.6: Hexagonales Abtastraster, das sich ganzzahlig auf kartesische Koordinaten abbilden läßt. Der Winkel der 1. charakteristischen Achse gegen die Horizontale beträgt $\arctan(2) = 63.4°$ statt $60°$. Gezeigt sind die Rezeptorpositionen für zwei verschiedene Auflösungsstufen mit Rezeptorabstand $d = 2$ (ausgefüllte Kreise) und $d = 4$ (große offene Kreise). Alle Rezeptorschichten sind um den Fixpunkt zentriert und verfügen über die gleiche Anzahl von Rezeptoren, so daß die Seitenlänge des jeweils erfaßten Bildbereichs proportional zum Rezeptorabstand ist. Rechts oben ist die Anordnung der Auflösungsschichten im Überblick dargestellt.

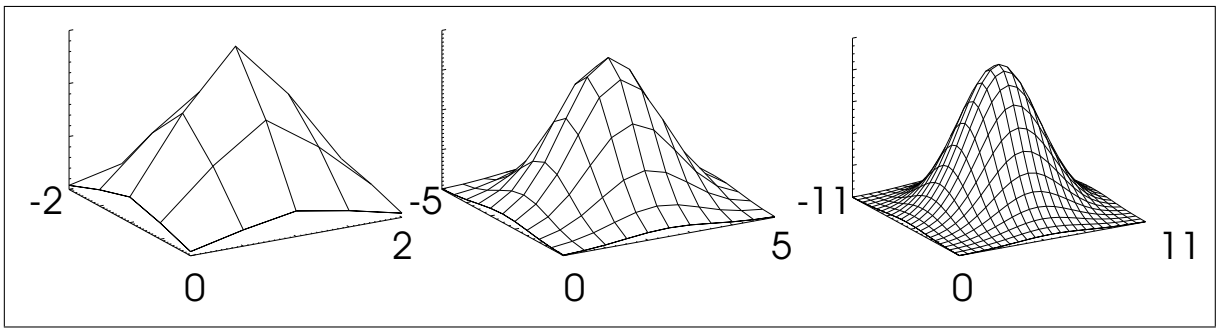


Abbildung 3.7: Zur Glättung des Eingangsbildes vor der Unterabtastung verwendete Gaußmasken. Dargestellt sind die Masken für die Abtastweiten $d = 2, 4$ und 8. Alle Masken sind auf Eins normiert, so daß die Gesamthelligkeit des Bildes bei Faltung mit der Maske nicht verändert wird.

Tabelle 3.1: Zeilen- bzw. Spaltenzahl M und Breite σ der verwendeten Gaußmasken bei verschiedenen Abtastweiten d (Daten aus [Spengler, 1996]).

d	**M × M**	σ
2	5×5	1.05
4	11×11	2.1
8	23×23	4.2

3.2.3.4 Modellierung der Neuronenschichten mit retinalen Ganglienzellen (X-Zellen)

Die Modellierung der in Abschnitt 3.2.1 besprochenen rezeptiven Felder der Ganglienzellen geschieht bei Abtastung mit dem pseudohexagonalen Gitter zweckmäßigerweise durch eine Linearkombination der RFs der Rezeptorzellen wie sie in Abb. 3.8 dargestellt ist. Die Ganglienzellen sind auf den gleichen Positionen wie die Rezeptoren angeordnet. Jede Ganglienzelle empfängt positiven Input vom Rezeptor am gleichen Ort und negativen Input von seinen sechs nächsten Nachbarn. Um eine ausgeglichene Summe zu erhalten, wird der zentrale Input sechsfach gegenüber dem der Nachbarn gewichtet. Dieses Modell zur Bildung einer Center-Surround-Verschaltung wird auch als *Difference of offset Gaussians* (DOOG) bezeichnet und hat sich als Standardverfahren in der technischen Bildverarbeitung etabliert [Hartmann, 1982; Young, 1986].

Formal entspricht dies wiederum einer Faltung mit einer diskreten Funktion des Bildortes, die die in Abb. 3.8 eingetragenen Gewichte als Werte erhält. Die beiden Faltungsoperationen lassen sich also zusammenfassen und ergeben so das effektive rezeptive Feld einer X-Zelle bei stationärer Reizung:

$$\begin{aligned} X(i,j) &= Z(i,j) - \frac{1}{6}\Big[\, Z(i+d,j) + Z(i-d,j) \\ &\quad + Z(i-\frac{d}{2}, j-d) + Z(i+\frac{d}{2}, j-d) \\ &\quad + Z(i-\frac{d}{2}, j+d) + Z(i+\frac{d}{2}, j+d)\,\Big] \end{aligned} \tag{3.7}$$

Wie man aus dem Vergleich von Abb. 3.8 und 3.2 erkennt, wird die Mexikanerhut-Funktion (die ja selbst nur eine Näherung darstellt) durch dieses Verfahren gut approximiert.

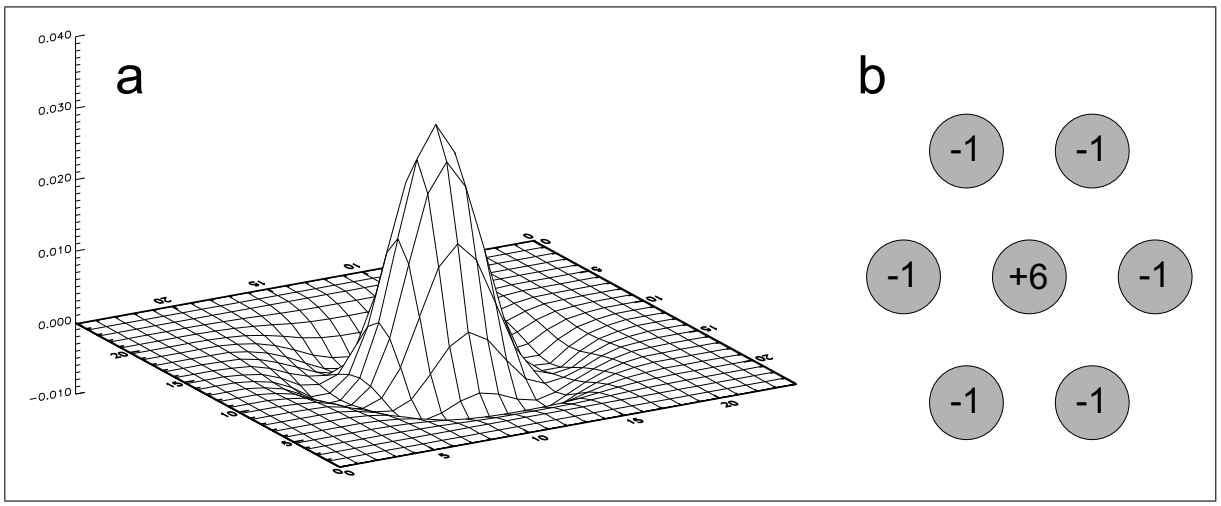


Abbildung 3.8: Rezeptives Feld $RF(i,j)$ der verwendeten X-ON-Zellen. Die Form entspricht recht gut derjenigen einer Mexikanerhut-Funktion aus Abb. 3.2. Das RF einer OFF-Zelle erhält man einfach durch Vorzeichenumkehr.

Die in Kap. 3.2.2 beschriebene getrennte zeitliche Tiefpaßfilterung der Eingangssignale aus Zentrum und Umfeld des RF liefert das gesamte dendritische Eingangssignal $G^{dendr}(t)$ einer Ganglienzelle. Um den Arbeitsbereich der Neurone zu vergrößern, d.h. einen größeren Kontrastbereich abdecken zu können, wird dieses Eingangssignal zur Berechnung des Membranpotentials $G(t)$ noch mit einer Sigmoidfunktion gestaucht:

$$G(t) = G^{max} \cdot \left[\frac{1}{1 + e^{-\alpha \cdot G^{dendr}(t)}} - \frac{1}{2} \right] \tag{3.8}$$

wobei α der Steigungsparameter der Sigmoiden ist.

Der dynamische Schwellenmechanismus und der Spike-Encoder aus Kap. 2.6 vervollständigen die bisher beschriebene Ganglienzelle zu einem impulscodierenden Modellneuron. Vom Marburger Modellneuron unterschiedet es sich nur durch die zusätzliche Anwendung der Sigmoidfunktion und das Fehlen der modulatorischen Linking-Eingänge.

Die von diesen Modellneuronen generierten Spikefolgen stellen das Ausgangssignal der Vorverarbeitung und gleichzeitig das Eingangssignal des Hardware-Accelerators dar (s. Kap. 3.2.4).

3.2.3.5 Wahl der geeigneten räumlichen Auflösung

Die Faltung mit der angenäherten Mexikanerhut-Funktion wirkt im Frequenzbereich als isotroper räumlicher Bandpaß. Das hat zur Folge, daß nur Kontrastgradienten erfaßt werden, deren Größenordnung in etwa mit dem Durchmesser des Mexikanerhuts übereinstimmt. Dies ist einerseits eine erwünschte Eigenschaft, da ja gerade bei der Kantendetektion die niederfrequenten Anteile, die große homogene Strukturen kennzeichnen, eliminiert werden sollen (räumlicher Hochpaß). Andererseits beschränkt die ebenfalls damit einhergehende Tiefpaßfilterung die mögliche Ortsauflösung und damit die Genauigkeit der internen Repräsentation. Selbstverständlich kann keine Filterfunktion eine Ortsauflösung herstellen, die besser als diejenige der Eingangsbilder ist. Liegen die Eingangsbilder vergleichsweise schlecht aufgelöst vor, so wird man also zur bestmöglichen Abtastung des Eingangssignals greifen; im oben vorgestellten Modell bedeutet das eine Abtastweite von 2 Pixel. Hat man es jedoch mit sehr gut aufgelösten Eingangsbildern zu tun, tritt das umgekehrte Problem auf: Die Anzahl der Neurone in jeder retinotopen Neuronenschicht des Systems hängt quadratisch von der Ortsauflösung, d.h. von der Abtastweite der Neuronenanordnung ab, so daß sich schnell für technische Zwecke inakzeptable Neuronenzahlen ergeben. Zudem werden bei gut aufgelösten Bildern häufig auch Merkmale miterfaßt, die für die gegebene Aufgabe irrelevant sind (etwa kleine Seitenstrukturen von Objekten). Hier bietet sich als geeignete Lösung die Vergrößerung der verwendeten Abtastweite, d.h. eine bewußte Verschlechterung der räumlichen Auflösung an.

Viele Anwendungen aus der praktischen Bildverarbeitung lassen sich in dieser Hinsicht eingrenzen und somit effizient realisieren. Insbesondere sind Anordungen, bei denen sich der Abstand der zu segmentierenden Objekte zur Kamera nicht oder nur geringfügig ändert, für eine feststehende räumliche Bandpaßfilterung geeignet. Im Gegensatz dazu erfordern alle Anwendungen, bei denen Kamera und Szenerie sich relativ zueinander bewegen, die Analyse eines großen Bereiches von Ortsfrequenzen. So durchläuft z.B. ein dem

Beobachter entgegenkommendes Auto auf einer geraden Landstraße alle Skalenbereiche von 'gerade noch aufzulösen' bis zur bildfüllenden Größe.

Das menschliche Sehsystem löst dieses Problem auf der Eingangsseite durch die stark inhomogene Anordung der Rezeptoren in der Retina (s. Kap. 2); zusätzlich sind intern offenbar mehrere Auflösungsstufen (also ein breites Band von Ortsfrequenzen) gleichzeitig repräsentiert. Diese können je nach Problemstellung zur Segmentierung bzw. Erkennung von Objekten herangezogen werden.

Im technischen Bereich hat sich als eine mögliche Nachbildung einer inhomogenen Bildabtastung die *Log-Polar-Transformation* Reitböck und Altmann [1984] etabliert. Dabei werden die Eingangsbilder zunächst in Polarkoordinaten umgerechnet, von dieser Darstellung dann der komplexe Logarithmus gebildet und zum Schluß das so verzerrte Bild wieder in kartesische Koordinaten zurücktransformiert. Während die erste Operation die Darstellung mit dem Abstand vom Bildzentrum als Koordinate ermöglicht, führt die Anwendung des Logarithums zu einer stark inhomogenen Darstellung der Bilder: Das um den Koordinatenursprung gelegene Bildzentrum wird gegenüber der Peripherie extrem vergrößert repräsentiert. Diese Art der Transformation hat für technische Anwendungen eine Reihe von Vorteilen:

- Auch bei feststehender Kamera kann ein großer Raumbereich erfaßt werden.
- gut aufgelöstes Bildzentrum
- einfache numerische Implementation (vgl. aber unten)

Dem stehen allerdings auch Nachteile gegenüber:

- Die Singularität des Logarithmus im Ursprung muß behoben oder umgangen werden.
- Die überall verzerrte Darstellung läßt keine einfachen räumlichen Vergleichsoperationen im ursprünglichen Koordinatensystem zu – hier steigt der numerische Aufwand enorm.

Insbesondere der letzte Punkt führt zu großen Problemen bei der Umsetzung der in Abschnitt 2.5.1 beschriebenen Gestaltgesetze für die Segmentierung: Gerade Linien werden durch die Log-Polar-Transformation nach außen gekrümmt, so daß eine einfache Zuordnung nicht mehr möglich ist. Interessanterweise verwendet zwar das menschliche Sehsystem in der Peripherie eine annähernd logarithmisch abnehmende Ortsauflösung, hat jedoch im zentralen Bereich der Fovea eine konstante Rezeptordichte. Muß ein größerer Bereich genau analysiert werden, so wird der Blick und damit die Fovea nacheinander auf Teilbereiche ausgerichtet.

Um einerseits entsprechend der in Kap. 1 erhobenen Forderung einen möglichst allgemeinen Ansatz zu wählen und andererseits das System technisch realisierbar zu halten, fiel die Wahl für das vorliegende Projekt auf die gleichzeitige Repräsentation der Szene auf mehreren Auflösungsstufen, deren Abtastweiten frei wählbar sind. Dabei verfügen alle Auflösungsstufen über gleich viele, um die Bildmitte zentrierte Neurone, so daß mit steigender Abtastweite automatisch der abgetastete Bereich proportional dazu vergrößert

wird. Durch Ausblenden der zentralen Bereiche der groben Auflösungsstufen läßt sich daraus eine stufenweise inhomogene Abtastung des Gesamtbildes gewinnen. Abb. 3.6 verdeutlicht die Anordnung der Rezeptoren bei zwei Auflösungsstufen.

3.2.4 Einbindung des Hardware-Accelerators in die technische Umgebung

Das in der dedizierten Hardware verwendete Modellneuron wurde bereits in Kap. 2.6.2 kurz vorgestellt. Verglichen mit anderen Verfahren der technischen Bildverarbeitung handelt es sich hier um einen relativ aufwendigen Codierungsansatz. Neben dem inhaltlichen Aspekt der biologienahen Modellierung bietet die vollständige Codierung mit impulscodierenden Neuronen jedoch auch Vorteile. Zum einen müssen die internen Zustandsvariablen der einzelnen Modellneurone außerhalb des jeweiligen Neurons nicht bekannt sein; die Ausgangsinformation erschöpft sich in der binären Codierung als 'Spike' oder 'kein Spike' zu jedem Zeitschritt. Zum zweiten sind unter normalen Umständen zu jedem Zeitpunkt immer nur wenige Neurone aktiv (unter 10% der Gesamtzahl). Dies führt dazu, daß trotz der aufwendigen inneren Verrechnung in jedem Zeitschritt nur wenige Spikes nach außen übertragen werden müssen. Eine effiziente Codierung besteht also darin, den gesamten Informationsstrom als Folge von Adressen bzw. Nummern aktiver Neurone darzustellen. Zur genauen Funktionsanalyse des Netzwerks (die ja ggf. auch interne Zustandsvariablen wie Membranpotentiale berücksichtigen muß) ist dann allerdings eine exakte Emulation der Acceleratorhardware durch ein konventionelles Programm erforderlich; diese wird durch die im Rahmen des Projektes entwickelten Programme *xsim* bzw. *msim* geleistet.

Weiterhin ist eine vom Simulationsalgorithmus unabhängige Beschreibung der Netzwerktopologie notwendig; diese wird erst zur Laufzeit erzeugt und in den Verbindungsspeicher des Accelerators geladen [Frank et al., 1996]. Ganz ähnlich wie bei der konventionellen Programmierung ist die direkte Binärdefinition in 'Maschinensprache' für eine funktionell orientierte Programmierung völlig ungeeignet; wünschenswert ist vielmehr ein einfach zu handhabendes Beschreibungsformat, das darüber hinaus auch einen unkomplizierten Austausch zwischen den am Projekt beteiligten Mitarbeitern ermöglichen sollte. Zu diesem Zweck wurde von Möller [1995] die Programmiersprache MNET entwickelt.

Ihre Syntax ist an die Sprache C angelehnt und erlaubt eine intuitive Beschreibung der Netzwerktopologie. Insbesondere erfolgt die Definition der Verschaltung zwischen den Neuronenschichten durch rechnerische Ausdrücke aus dem Befehlsvorrat von C, so daß die Verschaltung völlig frei programmierbar ist.

Ein Austausch einzelner funktioneller Module zwischen mehreren Programmierern ist ohne weiteres möglich und erlaubt ein effizientes gemeinsames Arbeiten an größeren Projekten wie dem hier vorgestellten.

Darüber hinaus ist MNET nicht auf den hier beschriebenen Accelerator bzw. seine Software-Emulation als Simulationsplattform festgelegt, sondern unterstützt andere Simulatoren wie z.B. den *Stuttgarter Neuronale Netze Simulator* (SNNS) [Zell, 1994].

3.3 Das Kontur-Form-System

Das Subsystem für die Verarbeitung stationärer Konturen (*Kontur-Form-System*) wurde von Weitzel [1998b] realisiert. Die im folgenden beschriebenen Feed-Forward- und lateralen Verschaltungen sind aus dieser Arbeit übernommen und werden hier lediglich zur besseren Verständlichkeit nochmals dargestellt. Die Analyse der Wechselwirkung zwischen Feeding- und Linking-Verbindungen in den Kapiteln 3.3.5 und 6 wurde von mir hinzugefügt und ergänzt die technische Implementation um theoretische Überlegungen.

3.3.1 Die Kantendetektoren

In Kap. 3.2 wurde vorgestellt, wie sich durch Bildung einer räumlichen Center-Surround-Verschaltung die Anwesenheit von Hell-Dunkel-Übergängen im rezeptiven Feld einer Zelle feststellen läßt. Allerdings war mit den dort verwendeten einfachen konzentrischen rezeptiven Feldern noch keine Unterscheidung bezüglich der Herkunft des Übergangs möglich: Eine Textur aus schwarzen Punkten auf hellem Grund spricht X-Zellen genauso (oder stärker) an wie eine über viele RFs hinweg verlaufende Kante. Wie schon in Kap. 6.3 angedeutet, wird eine große Klasse realer Gegenstände durch stückweise glatte Kanten begrenzt, die im Bild häufig entsprechend orientierte Helligkeitsgradienten bewirken. Aber auch viele unregelmäßig begrenzte Gegenstände wie z.B. Bäume erscheinen bei Betrachtung auf einer gröberen Auflösungsstufe mit einer glatten Kontur (vgl. auch Kap. 3.2.3.5). Die Entdeckung und gegenseitige Zuordnung solcher Objektkonturen ist die Aufgabe des Kontur-Form-Systems.

Die Detektion von Kanten, d.h. ausgedehnten, gerichteten Intensitätsgradienten, geschieht auf ähnliche Weise wie diejenige von ungerichteten: Mehrere Zellen mit einfachen RFs werden so auf eine Zielzelle verschaltet, daß diese ein RF mit den gewünschten räumlichen Eigenschaften erhält. In Kap. 6 wurde ein solches orientiertes RF bereits in einer vereinfachten Version zur Kantendetektion benutzt. Bei Verwendung der in diesem Kapitel vorgestellten X-Zellen läßt es sich auf einfache Weise aus mehreren X-ON und X-OFF-Zellen zusammensetzen. Abb. 3.9 zeigt das Prinzip der Verschaltung am Beispiel eines Kantendetektors aus jeweils drei X-ON- und X-OFF-Zellen.

Diese Form der Verschaltung wurde zum ersten Mal von Hubel und Wiesel [1962] für die Verschaltung vom CGL zu den orientierungssensitiven Zellen von V1 postuliert. Auf sie geht ebenfalls die Bezeichnung dieser Zellen als *Simple*-Zellen zurück.

Für einen sinnvollen praktischen Einsatz muß beim Entwurf der Verschaltung in zweierlei Hinsicht ein Kompromiß gefunden werden:

1. Die Parameter und Gewichte der beteiligten Neurone müssen so eingestellt sein, daß die Detektorzelle weder bereits bei kleinen, rauschbedingten Schwankungen der lokalen Grauwerte aktiv wird noch eine 'Schlagschatten'-Kante mit maximalem Kontrast zur Aktivierung benötigt. Derart ausgesprägte Kanten stellen in natürlichen Bildern die Ausnahme dar (s. auch Kap. 6.6.1).

2. Die Orientierungsselektivität der Zielzelle ist geeignet einzustellen. Je mehr X-Zellen auf eine Zielzelle aufgeschaltet werden, desto stärker wird deren Orientierungsselek-

tivität. Diese Eigenschaft ist zwar erwünscht, aber eine gewisse Unschärfe im Orientierungstuning ist notwendig, um auch Kanten detektieren zu können, die nicht genau einer der implementierten Vorzugsrichtungen entsprechen. Solche Zwischenorientierungen treten auch zwangsläufig bei gekrümmten Objektkonturen auf.

Insgesamt erweist sich der erste Punkt insofern als weniger kritisch, als die Zusammenführung mehrerer Neurone die Robustheit gegen Rauschen bereits erhöht. Im Fall eines normalverteilten, für jedes Neuron statistisch unabhängigen Rauschens sinkt dessen Amplitude mit der Quadratwurzel der Anzahl beteiligter Neurone. Bei anderen, räumlich korrelierten Rauschphänomenen, wie sie exemplarisch in Kap. 6.6 behandelt werden, ist der Effekt i.a. schwächer, aber dennoch vorhanden. Für den praktischen Einsatz entscheidend ist eine Anpassung der Parameter an die Gegebenheit der jeweiligen Aufgabenstellung, wobei der begrenzte Dynamikbereich der in Kap. 3.2.3 dargestellten Codierung bereits einen relativ engen Bereich für den Arbeitspunkt der Neurone bedingt. Voraussetzung dafür ist selbstverständlich, daß bereits bei der Bildaufnahme eine weitgehende Voranpassung erfolgt ist.

Der zweite Punkt läßt sich teilweise lösen, indem (wie in Abschnitt 3.2.3.5 beschrieben) mehrere Auflösungsstufen parallel betrachtet werden, die bei gleicher Neuronenanzahl verschiedene Ortsskalen im Eingangsbild analysieren. Dies hat zudem den Vorteil, daß die Verschaltungsdefinitionen von einer Auflösungsstufe zur anderen übernommen werden können – lediglich die Abtastweite d ändert sich.

3.3.2 Anordnung der Kantendetektoren für die verschiedenen Orientierungen

Im pseudohexagonalen Gitter bieten sich als Hauptorientierungen für die Kantendetektoren die Hauptachsen des Gitters an: $0°, 30°, 60°, 90°, \ldots 330°$. Eine ausführliche Diskussion

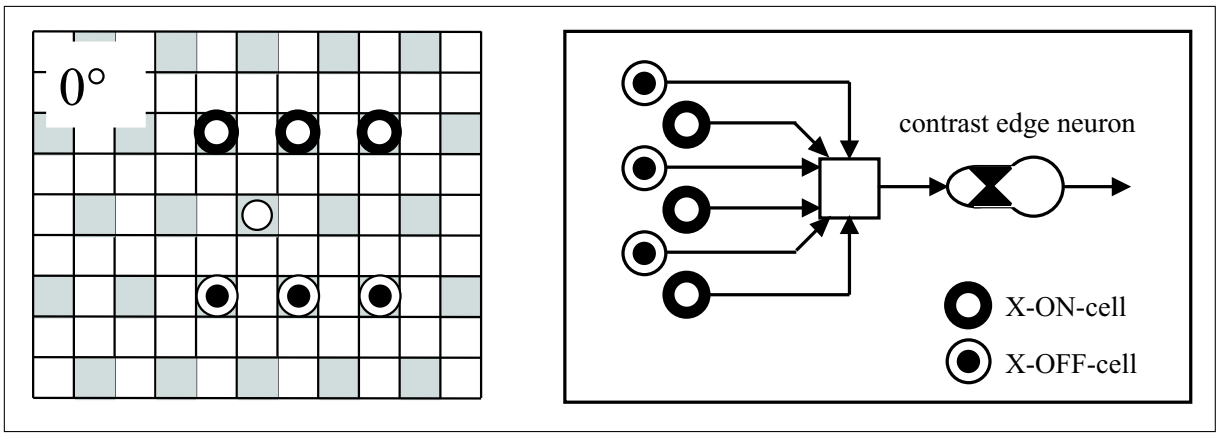


Abbildung 3.9: Kantendetektion am Beispiel eines 0°-Detektors. Die rezeptiven Felder von jeweils drei X-ON- und X-OFF-Zellen werden so überlagert, daß die Zielzelle speziell auf einen lokalen, gerichteten Helligkeitsgradienten anspricht. Je nach Einstellung der Parameter (insbesondere der statischen Schwelle Θ_0) kann auch eine teilweise Aktivierung der vorgeschalteten Zellen zur Anregung der Zielzelle genügen. Das Integral über das RF-Profil ist wie bei den X-Zellen Null; homogene Reize erregen die Zelle nicht.

der günstigsten Verschaltungsmuster für die verschiedenen Orientierungen unter Berücksichtigung der beim pseudohexagonalen Gitter auftretenden Besonderheiten findet sich bei Weitzel [1998b]. Aus den dort vorgestellten Varianten wurde die Kombination von ON- und OFF-Zellen wie sie in Abb. 3.9 dargestellt ist, übernommen. Weitzel diskutiert weitere, ausschließlich aus einer Sorte X-Zellen zusammengesetzte Typen von Kantendetektoren, die sich aber für die hier betrachteten Aufgaben als weniger vorteilhaft erweisen. Lediglich der $ON^2 - OFF^2$-Detektor, der eigentlich eine erweiterte Variante des ON-OFF-Detektors darstellt, ist diesem in einigen Belangen überlegen, erfordert jedoch die doppelte Anzahl an synaptischen Verbindungen, so daß er hier nicht verwendet wird. Aufgrund der vollständigen Formulierung in MNET (s. Kap. 3.2.4) war die Übernahme der Implementation von Weitzel ohne Probleme möglich. Abb. 3.10 zeigt die Verschaltungsmuster für die einzelnen Orientierungen. Die zugehörigen effektiven (stationären) rezeptiven Felder sind in Abb. 3.11 dargestellt.

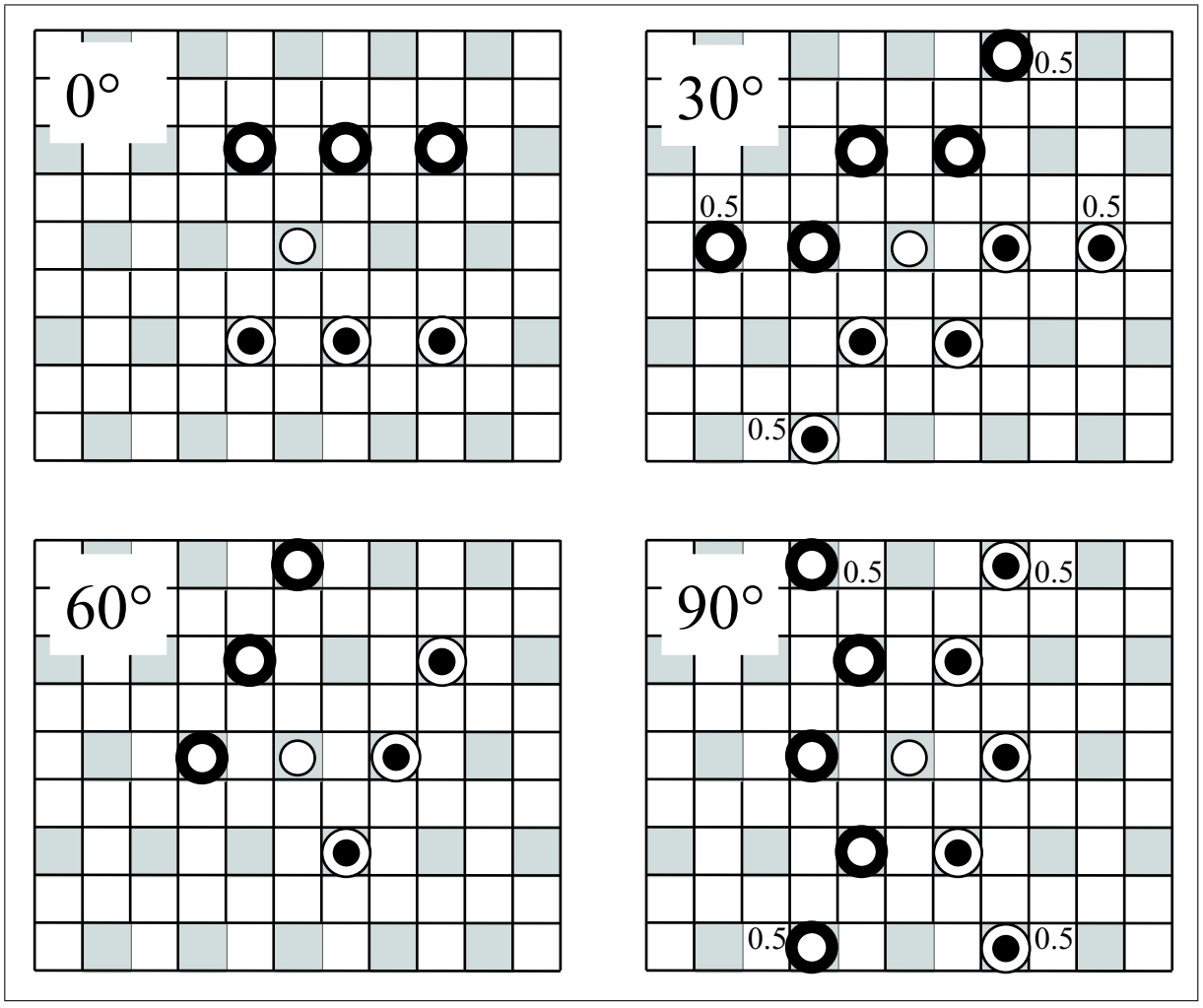


Abbildung 3.10: RF-Organisation der Kantendetektoren für die verschiedenen Vorzugsorientierungen im pseudohexagonalen Gitter. (Aus: [Weitzel, 1998b])

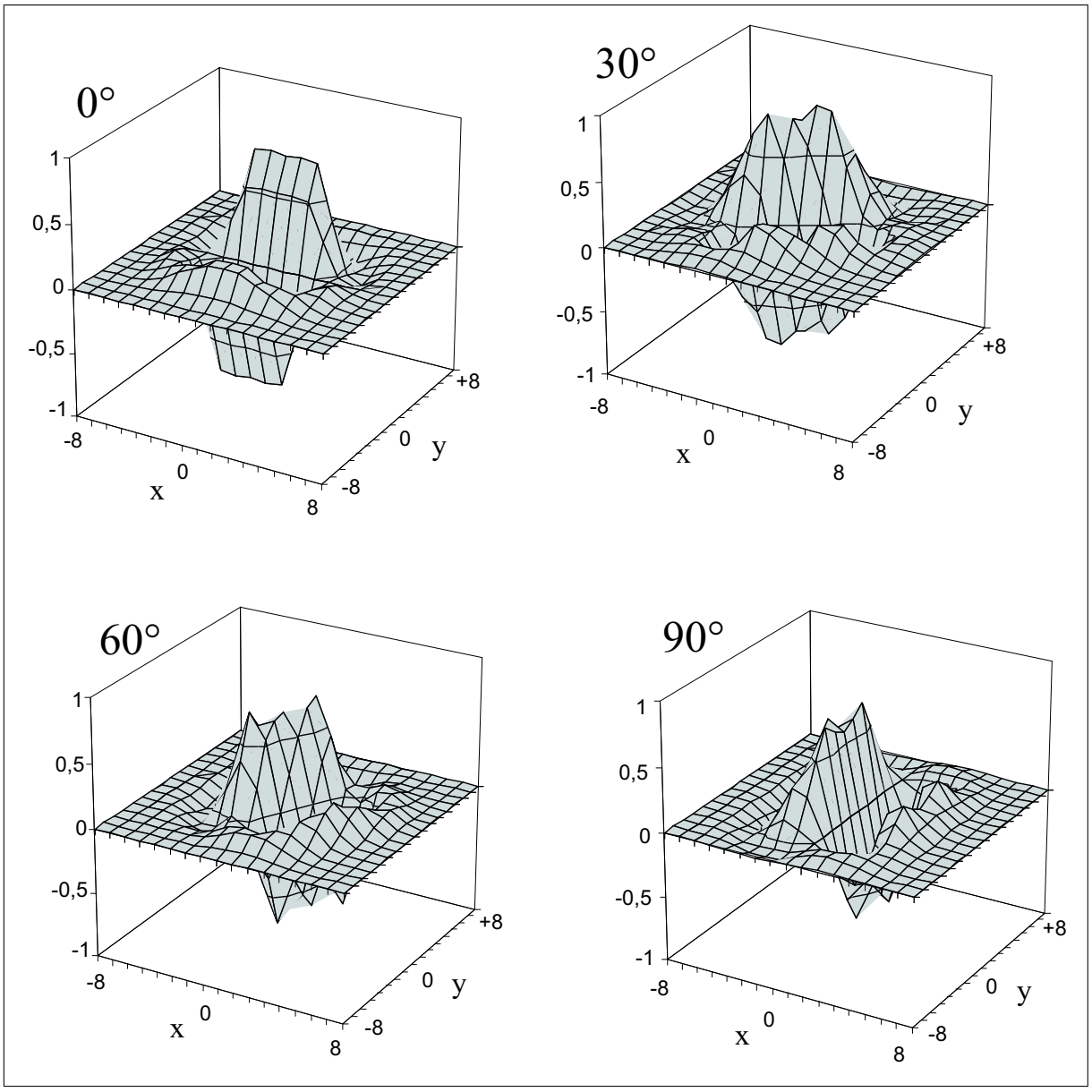


Abbildung 3.11: Effektive rezeptive Felder der Liniendetektoren für 0°, 30°, 60° und 90°. Die weiteren Orientierungen ergeben sich durch einfache Spiegelungen. (Aus: [Weitzel, 1998b])

3.3.3 Die laterale Linking-Verschaltung

3.3.3.1 Räumliches Profil des Linking

Das Kontur-Form-System hat die Aufgabe, die mit den im vorigen Abschnitt beschriebenen Kantendetektoren gefundenen Objektkonturen einander zuzuordnen und die Zuordnung bzw. Trennung entsprechend der Synchronisationshypothese im Zeitverlauf seiner Aktivität zu codieren.

Die in Kap. 2.5.1 vorgestellten Gestaltgesetze über die Wahrnehmung durchlaufender Linienzüge können direkt in die neuronale Verschaltung der Liniendetektoren einmodelliert werden [Eckhorn et al., 1990]. Dazu werden Kantendetektoren mit annähernd

kollinear angeordneten rezeptiven Feldern exzitatorisch über die Linking-Synapsen gekoppelt. Die räumlichen und zeitlichen Aspekte der Kopplung werden in den folgenden Abschnitten genauer vorgestellt und diskutiert.

In Abb. 3.12 ist die genaue Anordnung der Linking-Nachbarschaft eines Kantendetektors im pseudohexagonalen Raster dargestellt. Die Kopplung ist symmetrisch ausgeführt; nicht exakt koaxial orientierte Nachbarn sind mit schwächeren Gewichten verbunden.

3.3.4 Auswirkungen der Linking-Architektur

Aus den in Kap. 2.5.1 vorgestellten Überlegungen (die in Kap. 6 noch präzisiert werden) läßt sich die Feststellung ableiten, daß Konturen – zumindest in natürlichen Bildern – fast nie isoliert auftreten, sondern gerade zusammengehörige Kontursegmente sich erstens durch *räumliche Nähe* und zweitens durch *kollineare Orientierung* auszeichnen. Umgekehrt läßt sich aus solchen Eigenschaften auf die Zusammengehörigkeit bzw. Trennung von Kontursegmenten zurückschließen. Diese Tatsache läßt sich in zweifacher Hinsicht nutzbar machen:

1. **Räumlicher Aspekt:** Durch exzitatorische Kopplung zwischen benachbarten, kollinear orientierten Kontursegmenten wird die typische Korrelationsstruktur natürlicher Bilder unterstützt (s. auch Kap. 6).

2. **Zeitlicher Codierungsaspekt:** Soll entsprechend der Synchronisationshypothese die Zusammenghörigkeit von Kontursegmenten durch synchrone Aktivität codiert werden, so läßt sich dies durch eine geeignete Wahl der Kopplungsparameter in Punkt 1 bewerkstelligen.

Der erste Punkt stützt sich weitgehend auf die Untersuchungen aus Kap. 6 und wird dort ausführlich diskutiert. Der zeitliche Aspekt wird in Kap. 4.1 in seiner Beziehung zur Segmentierungsdynamik behandelt. An die lateralen Kopplungsverbindungen sind in Bezug auf die zeitlichen Eigenschaften zunächst drei Anforderungen zu stellen:

1. Sie sollen die Aktivität räumlich zusammengehöriger Kontursegmente synchronisieren. Im Spezialfall der stationären Reizung bedeutet das eine Angleichung der Phasen während der (quasi-)stationären Oszillation (s. auch Kap. 4.1).

2. Die stationäre Feuerrate der gekoppelten Neuronen soll möglichst wenig verändert werden.

3. Die räumlichen RF-Eigenschaften der Kantendetektoren sollen nur insofern verändert werden, daß die *bedingte Antwortwahrscheinlichkeit* eines Detektors erhöht wird, wenn bereits kollineare Nachbarkonturen von anderen Neuronen angezeigt werden.

Die erste Anforderung läßt sich leicht durch eine exzitatorische Kopplung mit minimaler Verzögerung (1 bin) erfüllen. Jedes aktive Konturneuron erregt die mit ihm verbundenen Nachbarn zusätzlich und bringt sie somit näher an die Feuerschwelle. Haben die betreffenden Neurone bereits ähnliche Phasen (z.B. weil sie seit Beginn der Reizung

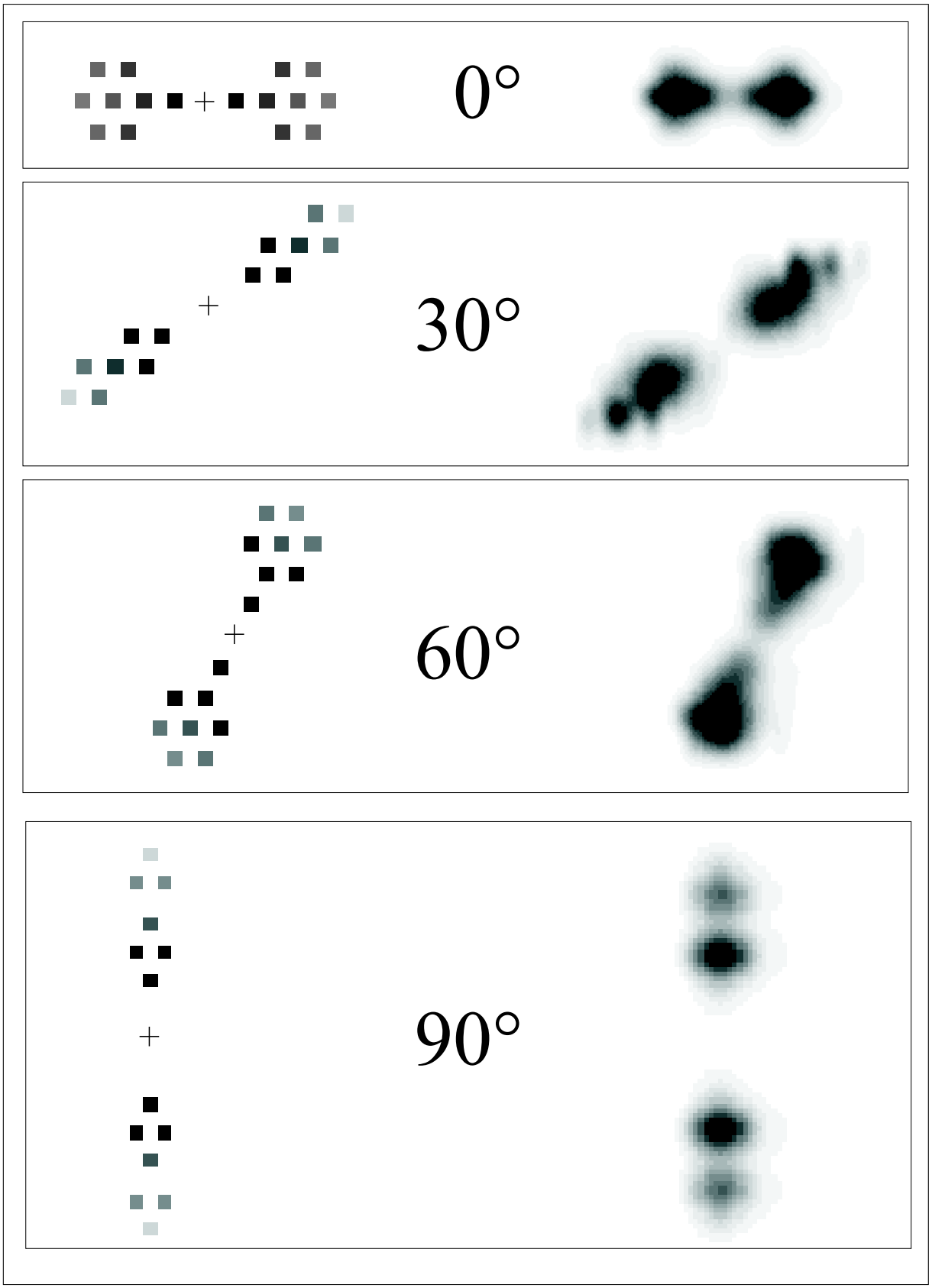


Abbildung 3.12: Räumliche Verteilung der Linking-Gewichte für die verschiedenen Hauptorientierungen im pseudohexagonalen Gitter, exemplarisch für ein Neuron. Der Ort des Neurons ist mit + bezeichnet. Links sind die Gewichte als Grauwerte auf den diskreten Gitterposition dargestellt; die rechte Darstellung berücksichtigt die Glättung in der Vorverarbeitung. Die Anordnung für die weiteren Orientierungen ergibt sich durch Spiegelung an der vertikalen bzw. horizontalen Achse.

ähnlich angeregt wurden), so läßt sich durch eine ausreichend große Amplitude des koppelnden EPSP praktisch sicherstellen, daß das in der Phase etwas zurückliegende Neuron durch das koppelnde EPSP sofort überschwellig wird, also mit nur einem Zeitschritt Versatz zum erregenden (ihm vorlaufenden) Neuron feuert. Diese schnelle Wirkung der Kopplung ist eine entscheidende Voraussetzung für die erfolgreiche Herausbildung einer Segmentierungsdynamik, wie sie in Kap. 4.1 beschrieben wird.

Eine hohe Amplitude des koppelnden EPSP steht andererseits im Widerspruch zur zweiten Forderung, nach der die stationäre Feuerrate und damit das Membranpotential der Kantendetektoren durch die laterale Kopplung gegenüber dem ungekoppelten Fall möglichst nicht verändert werden sollen. Dieser Widerspruch läßt sich allerdings lösen, wenn das koppelnde EPSP nicht nur eine große Amplitude, sondern zugleich auch eine kurze Zeitkonstante aufweist. In diesem Fall wird zwar das Membranpotential nach dem Eintreffen des auslösenden Kopplungsspikes kurzzeitig stark erhöht (was ja im Sinn einer schnellen Wirkung der Kopplung erwünscht ist). Aufgrund der kurzen Zeitkonstante der Kopplungssynapse klingt die Wirkung jedoch schnell wieder ab. Liegt die Zeitkonstante der Kopplung deutlich unter der Oszillationsperiode der Neurone, so kann ihre Wirkung nach einer Periode (wenn das Membranpotential wieder in der Größenordnung der Schwelle ist) in guter Näherung vernachlässigt werden. Der Zeitpunkt der nächsten Spikeauslösung wird dann allein vom direkten Input aus dem RF sowie eventuell neu erzeugten Kopplungsspikes bestimmt.

3.3.5 Interaktion von Linking und Feeding

Die von Eckhorn et al. [1990] vorgeschlagenen lateralen Koppelverbindungen unterstützen die Detektion sowie die Synchronisation von Aktivität längs durchlaufender Konturen in spezifischer Weise. In diesem Abschnitt soll der Einfluß der Kopplungsverbindungen auf die Neuronenantwort genauer dargestellt und insbesondere die Unterschiede zwischen additiver und multiplikativer Kopplung hinsichtlich eines 'effektiven RFs' der Zellen untersucht werden.

Die Bezeichungen entsprechen denen aus Kap. 2. Nach Gl. 2.2 ist das Membranpotential eines Neurons bei einer additiven bzw. multiplikativen Nachbarschaftskopplung zu jedem Zeitpunkt $t$gegeben durch

$$U^{(a)}(t) = F(t) + w^{(a)}W^{(a)}(t) \qquad \text{und} \qquad U^{(m)}(t) = F(t) \cdot (1 + w^{(m)}W^{(m)}(t)) \tag{3.9}$$

Diese Gleichungen beschreiben die Kennflächen des dendritischen Neuronenbereichs über dem von $x(t)$ und $W(t)$ aufgespannten zweidimensionalen Eingangsraum aus 'originärem' Input aus dem RF und dem Kopplungsbeitrag. In Abb. 3.13 erkennt man deutlich, daß die multiplikative Kopplung ohne Aktivierung des eigentlichen RF keine Antwort hervorrufen kann.

Der einfacheren Bezeichnung halber setzen wir für die folgenden Betrachungen $V_F = 1$, d.h. wir messen alle Größen in Einheiten der Feeding-Verstärkung. Damit bekommt das Verhältnis von Kopplungs- zu Feedingverstärkung die Rolle eines allgemeinen Kopplungsparameters. Kap. 6. Im gemeinsamen Grenzfall $w = 0$ (keine Kopplung) verhalten sich

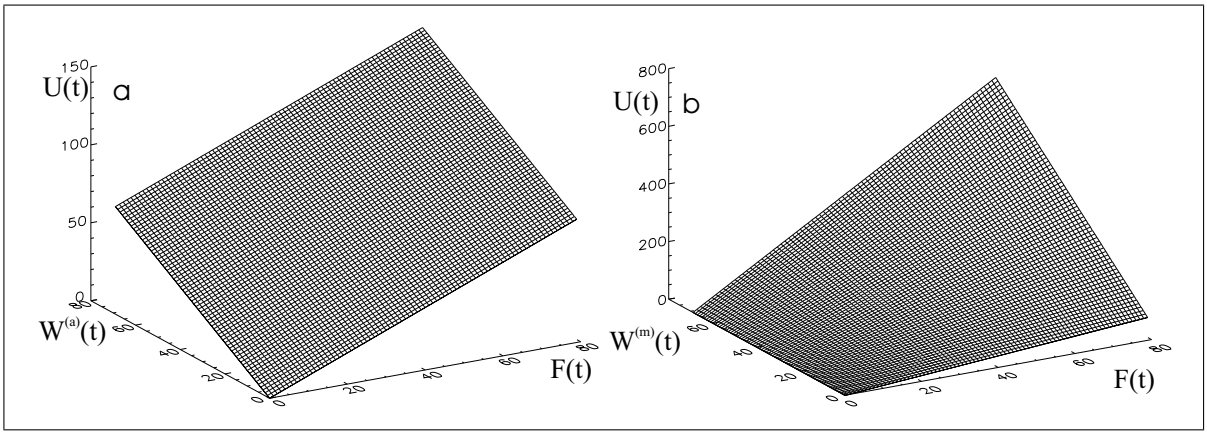


Abbildung 3.13: Wechselwirkung von Feeding-Input und additivem (a) bzw. multiplikativem (b) Kopplungsinput bei der Entstehung des Membranpotentials am Eingang des Spike-Encoders.

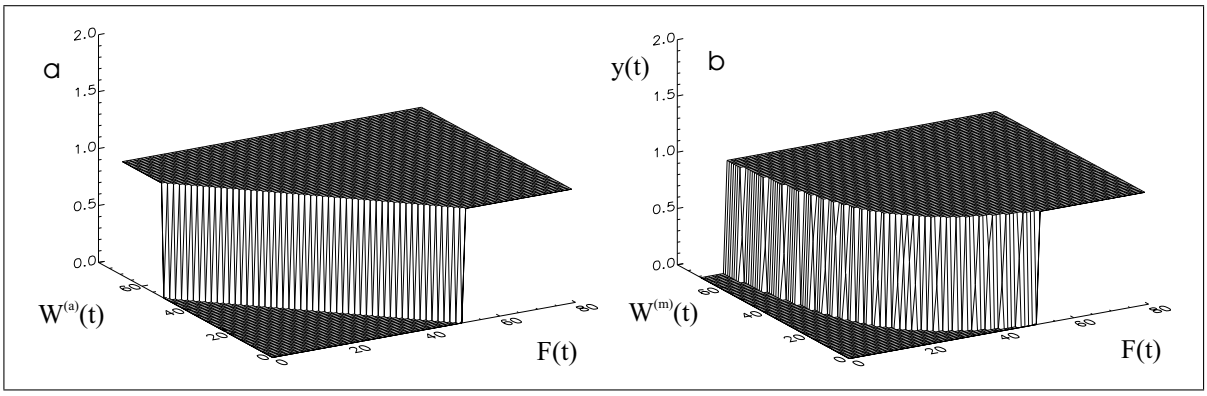


Abbildung 3.14: Ausgangscharakteristik des Modellneurons bei gleichzeitigem Feeding- und Kopplungsinput. Die Abbildung entspricht den durch Schwellenvergleich binarisierten Kennflächen aus Abb. 3.13.

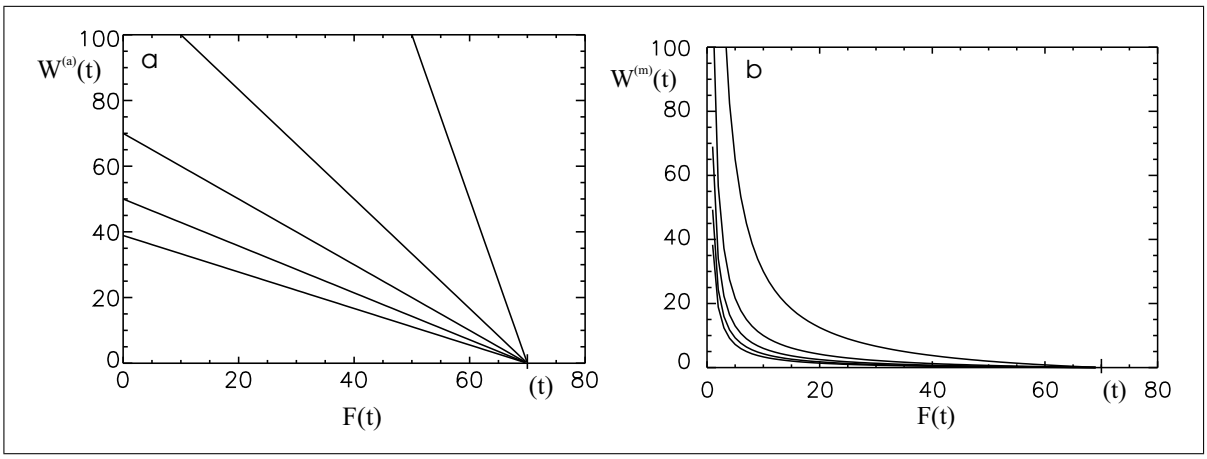


Abbildung 3.15: Begrenzung der über- und unterschwelligen Bereiche des Eingangsraums aus Abb. 3.14 für verschiedene Werte der Kopplungsstärke w. (a) Im Fall der additiven Kopplung ergibt sich eine Schar von Geraden, deren Achsenabschnitt $\Theta/w^{(a)}$ mit steigender Kopplungsstärke sinkt. (b) Für die multiplikative Kopplung ergibt sich als Grenzen eine Schar von Hyperbeln, die für größeres $w^{(m)}$ stärker 'durchgebogen' sind. Alle Kurven schneiden die Abszisse bei $F(t) = \Theta(t)$; dies entspricht genau der Schwellwertbedingung im ungekoppelten Fall ($w = 0$).

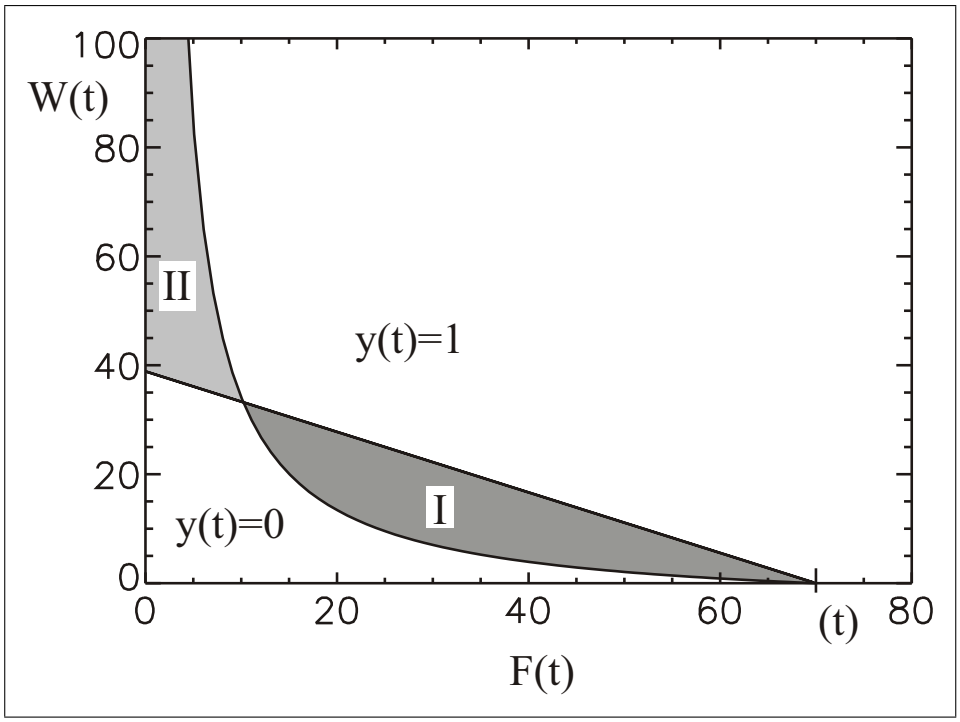


Abbildung 3.16: Ein direkter Vergleich der Grenzlinien im Eingangsraum verdeutlicht die Bereiche, in denen additive und multiplikative Nachbarschaftskopplung sich in Bezug auf die Antwort der Neurone unterscheiden. Bereich I entspricht einem schwellennahen Membranpotential; das Neuron wird bei multiplikativer Kopplung etwas früher überschwellig als bei additiver. Bereich II beschreibt eine Situation wie sie für Linienenden bzw. Konturecken typisch ist; die mulitplikative Kopplung muß hier viel stärker sein, damit das Neuron überschwellig werden kann. Im verrauschten Fall verwischen diese Grenzen allerdings (s. Kap. 6.4.3).

die additive und mulitplikative Kopplungsvariante selbstverständlich gleich; ein Aktionspotential wird hier einfach für $x(t) > \Theta(t)$ ausgelöst.

Die lokalen Kennliniensteilheiten des Membranpotentials in Bezug auf den originären Feeding-Input sind:

$$\frac{\partial U^{(a)}}{\partial F} = 1 \qquad \text{und} \qquad \frac{\partial U^{(m)}}{\partial F} = 1 + w^{(m)} W^{(m)}(t) \tag{3.10}$$

Im additiven Fall ist die Steigung dieser Kennlinie also konstant, während sie im multiplikativen Fall von der momentanen Größe des Kopplungsbeitrags abhängt.

Die Bedingung für die Auslösung eines Spikes läßt sich zur Charakterisierung der beiden Kopplungsvarianten heranziehen: Die Grenzlinie der in Abb. 3.14 dargestellten binarisierten Ausgangskennfläche ist in Abb. 3.15 für verschiedene Werte von $w^{(a)}$ und $w^{(m)}$ eingezeichnet. Abb. 3.16 zeigt exemplarisch den Verlauf der Grenzen für je einen Wert von $w^{(a)}$ und $w^{(m)}$.

In denjenigen Bereichen im Eingangsraum, die unter- bzw. oberhalb *beider* Kurven liegen, haben beide Kopplungsvarianten die gleiche Auswirkung. Unterschiede zeigen sich in komplementärer Weise in zwei Bereichen:

- Bereich I entspricht einem schwellennahen Arbeitspunkt des Neurons. Die multiplikative Kopplung macht hier bei gleich starkem Input das Neuron früher überschwel-

lig als ihr additives Gegenstück. Dieser Effekt ist um so ausgeprägter, je größer $w^{(m)}$ ist.

- Bereich II hingegen repräsentiert eine Situation mit weit unterschwelligem Feeding-, aber starkem Kopplungsinput. Aufgrund des hohen Anteils zusammenhängender Konturen in natürlicher Bildern (s. Kap. 6.6) ist dies von der reinen Häufigkeit her eine eher unbedeutende Merkmalskombination. Die Wichtigkeit dieser Region des Eingangsraums wird erst klar, wenn man bedenkt, daß sie genau die Situation eines Neurons am Ende einer Kontur bzw. Linie charakterisiert. Tatsächlich kann eine additive Kopplung hier trotz fehlenden Feeding-Inputs das betreffende Neuron überschwellig werden lassen. Bei multiplikativer Kopplung ist das (zumindest im unverrauschten Fall) nicht möglich.

3.4 Das Transientensystem

Zur Verarbeitung und Analyse bewegter Bilder ist eine Verarbeitung wie sie in Kap. 3.3 vorgestellt wurde, nur bedingt geeignet. Die Kantendetektion gelingt um so besser, je länger der Input zur Verfügung steht, da zeitliche Rauschphänomene ausgemittelt werden können. Räumliche Verwacklungen ('Jitter') können nur in der Größenordnung der RFs der Rezeptoren bzw. Kantendetektoren ausgeglichen werden. Zwar kann die Einführung von *Complex-Zellen* mit einer gewissen Ortsinvarianz, wie sie z.B. von Hartmann [1982] vorgeschlagen wurde, den räumlichen 'Fangbereich' von Kantendetektoren vergrößern, aber auch hier ist eine zeitliche Integration erforderlich, um Objektkonturen zuverlässig zu detektieren. Bewegte Objekte in einer ansonsten stationären Szene können also nur dann mit stationären Kantendetektoren registriert (und ggf. segmentiert) werden, wenn sie sich so langsam bewegen, daß die Zeit, die sie zum Durchqueren des RF eines Kantendetektors (bzw. einer Complex-Zelle) benötigen, größer oder gleich der für eine erste Antwort benötigten Integrationszeit des betreffenden Detektors ist. Die Herausbildung einer Segmentierungsdynamik wie sie in Kap. 4.1 beschrieben wird, benötigt i.a. ebenfalls stationäre Verhältnisse.

Umgekehrt ist die Isolation und damit die Segmentierung bewegter Objekte vergleichsweise leicht möglich, wenn ein geeigneter Verarbeitungszweig mit durchgängig schnellen Antworteigenschaften in der Lage ist, transiente Anteile des Eingangssignals zu verarbeiten und auf ihre Signifikanz und Zusammengehörigkeit hin zu analysieren. Dies ist die Aufgabe des Transientensystems.

Eine zweite Funktion kommt hinzu, wenn man mit einer bewegten Kamera arbeitet. Aus der dynamischen Szene müssen zunächst die im Sinne einer Segmentierung 'interessanten' Bereiche (d.h. Kandidaten für sich bewegende Objekte) extrahiert werden, bevor durch eine Blickbewegung der 'interessanteste' Objektkandidat zur genaueren Analyse in den zentralen Sehbereich gebracht werden kann. Schließlich muß bei der Analyse eines bewegten Objekts die Kamera nachgeführt werden (Folgebewegung), um ein annähernd stationäres Bild auf der Retina zu ermöglichen. Das ist zwar Aufgabe der in Kap. 3.5.5 beschriebenen Aufmerksamkeitssteuerung, aber diese benötigt zuverlässige Eingangsinformation über den momentanen Aufenthaltsort und den Bewegungsvektor des verfolgten Objekts. Zudem führen während einer solchen Kamerafahrt große Bildbereiche Scheinbewegungen aus; trotzdem sollen Objektort und -geschwindigkeit richtig detektiert werden.

Ein Modell zur Bewegungs- und Transientenverarbeitung, das diese Anforderungen erfüllt, wurde von Schott [1999] vorgestellt und mit nur geringen Anpassungen für die vorliegende Arbeit übernommen. Wie beim Kontur-Form-System war dies aufgrund der vollständigen Formulierung in MNET ohne Probleme möglich. Die vollständige Darstellung ist in der Arbeit von Schott zu finden; hier kann aus Platzgründen nur eine zusammenfassende Darstellung erfolgen.

3.4.1 Aufbau des Transientensystems

3.4.1.1 Arbeitsweise und Eigenschaften der Bewegungsdetektoren

Entsprechend den zeitlichen Eigenschaften der Y-Ganglienzellen (s. Kap. 3.2) besteht die erste Stufe des Transientensystems aus der Detektion transienter Signalanteile in den Eingangsbildern, also lokaler zeitlicher Grauwertänderungen. Dies entspricht qualitativ der Bildung einer partiellen Zeitableitung an jedem Punkt der Eingangsbildfolge. Da jede Bewegung von Bildteilen mit solchen zeitlichen Änderungen verbunden ist, kann das Ergebnis dieser Transientendetektion zur Bewegungsanalyse herangezogen werden. Umgekehrt zeigt allerdings nicht jedes transiente Ereignis auch eine Bewegung an. Beispielsweise ist der eingeschaltete Blinker eines stehenden Autos eine ständige Quelle schneller Helligkeitsänderungen, ohne daß eine Bewegung stattfindet. Für eine Bewegungsanalyse ist die Einbeziehung der räumlichen Komponente zwingend erforderlich. Ein einfaches Modell für einen neuronalen Mechanismus, der eine Bewegungsschätzung auf der Grundlage raumzeitlicher Integration leistet, wurde 1957 von Reichardt gegeben. Seine Arbeitsweise ist in Abb. 3.17 dargestellt.

Der Reichardt-Detektor nutzt die Tatsache aus, daß eine Kontrastkante, die sich über den Sichtbereich bewegt, nacheinander an in Bewegungsrichtung gegeneinander versetzten Punkten eine gleichsinnige Änderung der lokalen Helligkeit hervorruft. Der Quotient aus zeitlichem und räumlichem Abstand kann dann zur Schätzung der Geschwindigkeit herangezogen werden.

$$v_{est}^{Reichardt} = \frac{\Delta x}{\Delta t} \tag{3.11}$$

Es ist klar, daß für derartige bilokale Detektoren die oben erwähnte Einschränkung gilt, daß zwar bewegte Kontrastkanten immer systematische transiente Ereignisse hervorrufen, die Beobachtung solcher Ereignisse jedoch keinen eindeutigen Rückschluß auf eine Bewegung im Bild erlaubt. Beispielsweise kann eine Kontrastkante an einem Ort verschwinden und nach der Zeit Δt an einem um Δx versetzten Ort eine andere, gleichartige Kante auftauchen. Für den bilokal arbeitenden Bewegungsdetektor besteht keine Möglichkeit, diesen Vorgang von einer echten Bewegung zu unterscheiden. Diese Einschränkung ist prinzipieller Natur und nur durch Einbeziehung weiterer Informationen (etwa von benachbarten Detektoren) aufzulösen. Bezeichnenderweise unterliegen auch menschliche Beobachter dieser Täuschung; auch ein Mensch hat ohne zuätzliche Information keine Möglichkeit, eine solche scheinbare Bewegung (*apparent motion*) von einer echten zu unterscheiden.

Das hier angesprochene Problem wird auch als *Korrespondenzproblem* bezeichnet, da es von der Unklarheit herrührt, welches Pixel in einem Kamerabild mit welchem anderen Pixel im folgenden Kamerabild korrespondiert. Die Pixel zweier aufeinanderfolgender Kamerabilder sind ununterscheidbar; es kann lediglich eine Schätzung erfolgen.

Das Korrespondenzproblem an sich tritt bereits bei der eindimensionalen Bewegungsschätzung auf, d.h. wenn bei bekannter Bewegungsrichtung lediglich der Betrag der Geschwindigkeit zu schätzen ist. Beim Versuch, mit lokal arbeitenden Detektoren Bewegung in zwei Dimensionen aus Bildfolgen zu extrahieren, stößt man auf eine weitere Ausprägung des Korrespondenzproblems, die als *Aperturproblem* bekannt ist [Wallach, 1935]: Be-

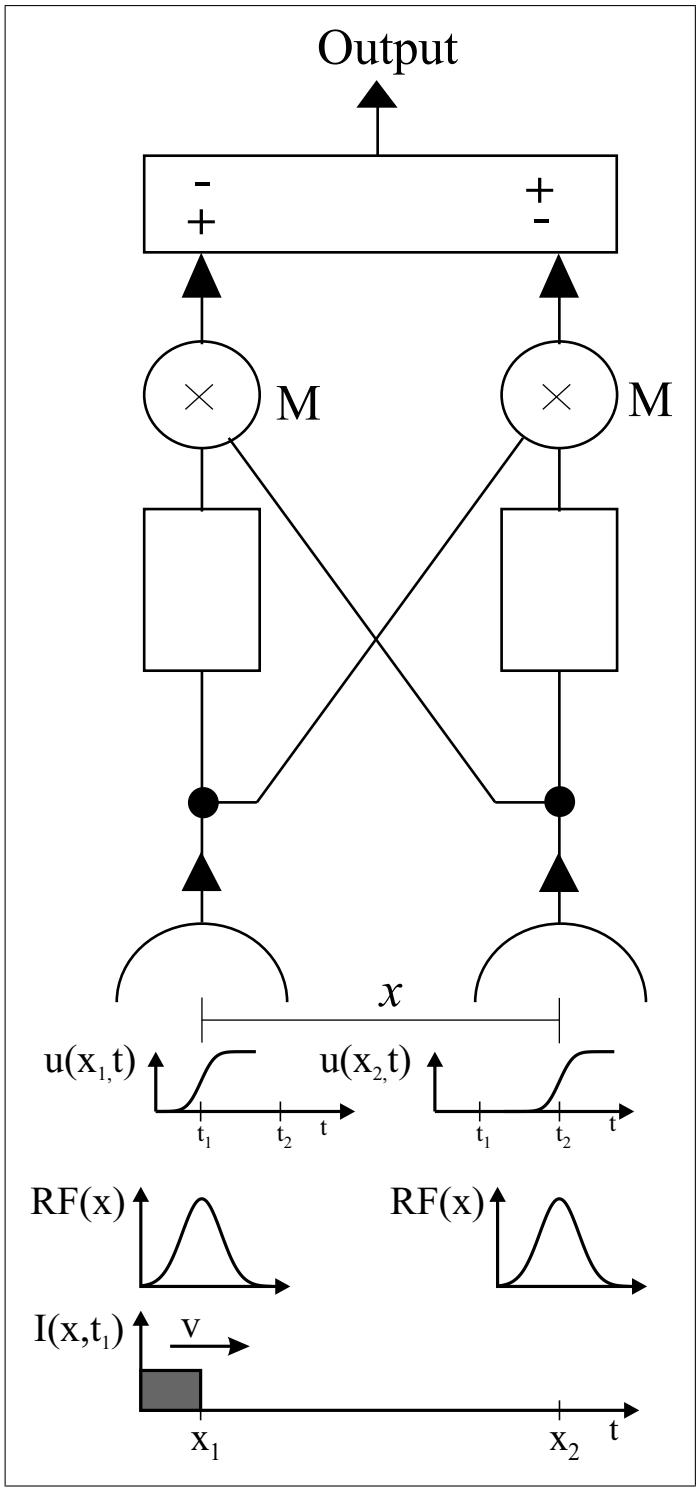


Abbildung 3.17: : Arbeitsweise des Reichardt-Detektors. Eine als Helligkeitsprofil $I(x)$ dargestellte Kontrastkante, die sich mit der Geschwindigkeit v bewegt, überstreicht in der Zeit Δt zwei Neurone mit dem räumlichen Abstand Δx. Nach der Faltung mit dem RF der Neurone wird deren Ausgangsaktivität mit einer Verzögerung Δ mit der Antwort des jeweils anderen Neurons multipliziert. Die beiden Verzögerungsrichtungen entsprechen zwei entegegengesetzten Bewegungsrichtungen; die Ergebnisse der beiden Multiplikationen werden deshalb noch subtrahiert, um die Gesamtantwort des Detektors zu erhalten. Die stärkste positive Anwort entsteht, wenn eine in positiver x-Richtung bewegte Kante genau die Zeit Δt benötigt, um beide Neurone zu überstreichen, also $\Delta = \Delta t$ ist. Je nach Ausdehnung der beteiligten RFs wird auch eine Antwort hervorgerufen, wenn dieses Kriterium nicht exakt erfüllt ist; diese ist dann allerdings schwächer. Im Endeffekt entsteht so ein Detektor mit einem kontinuierlichen Tuning für den Betrag der Geschwindigkeit in x-Richtung. (Nach: [Reichardt, 1957])

obachtet man eine sich bewegende gerade Kante durch eine runde Apertur, so ist die Bewegungsrichtung nicht eindeutig festzustellen; die wahrgenommene Bewegungsrichtung kann in um bis zu 90° nach oben oder unten von der tatsächlichen Richtung abweichen (s. Abb. 3.18).

Das visuelle System des Menschen erzeugt hier allerdings immer eine eindeutige Darstellung: Die Bewegungsrichtung wird als rechtwinklig zur Kante wahrgenommen.

3.4.1.2 Möglichkeiten zur Lösung des Korrespondenzproblems

Wie bereits angedeutet, läßt sich das Korrespondenzproblem nicht im strengen Sinn lösen. Allerdings kann die Wahrscheinlichkeit falscher Bewegungsschätzungen durch die Einbeziehung zusätzlicher Informationen stark reduziert werden. Dabei ist grundsätzlich zu unterscheiden, ob die zusätzliche Information lediglich aus den bisher diskutierten frühen visuellen Verarbeitungsstufen kommt, oder ob Objektwissen aus höheren Stufen der Verarbeitung herangezogen werden kann. Letzere Strategie spielt bei der Bewegungsschätzung durch menschliche Beobachter mit Sicherheit eine große Rolle; für die vorliegende Arbeit kommt jedoch nur die Auswertung von Low-Level-Informationen in Frage. Ähnlich wie bei der Implementation der Gestaltgesetze für das Zusammenbinden von Konturen werden dabei möglichst wenige, grundlegende Annahmen über die zu detektierenden Objekte zugrundegelegt und implementiert. Im Fall der Bewegungsschätzung ist dies in erster Linie die *Trägheitsannahme* und die *Annahme von der Objektkonstanz*.

Die Trägheitsannahme besagt, daß physikalische Objekte aufgrund ihrer Massenträgheit ihre Bewegungsrichtung und -geschwindigkeit nur selten abrupt ändern, sondern daß diese Änderungen i.a. stetig erfolgen. Voraussetzung für eine erfolgreiche Anwendung dieses Prinzips ist selbstverständlich eine ausreichend schnelle, an die jeweilige dynamische Szene angepaßte, zeitliche Abtastung der Eingangsbilder.

Die Annahme von der Objektkonstanz ist die Grundlage für die gegenseitige Zuord-

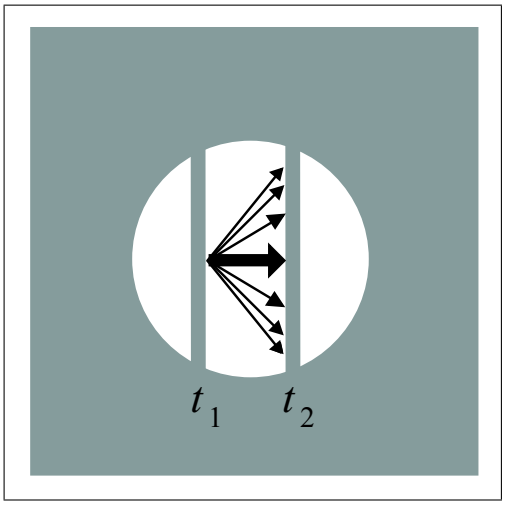


Abbildung 3.18: Das Aperturproblem. Wird die Beobachtung einer bewegten Kante durch eine Apertur eingeschränkt, so ist eine zweifelsfreie Zuordnung der Pixel durch eine übergeordnete Objektwahrnehmung nicht mehr möglich. Ein menschlicher Beobachter sieht jedoch immer eine Bewegung senkrecht zur Kante. (Aus: [Schott, 1999]).

nung (Korrespondenz) gleichartiger Pixel in aufeinanderfolgenden Eingangsbildern: Man geht davon aus, daß Objekteigenschaften (hier der Kantenkontrast) sich zwischen zwei Kamerabildern nur so wenig ändern, daß der Bewegungsdetektor die beiden aufeinanderfolgenden Signale als gleichartig erkennt. Ist dies nicht gewährleistet (etwa wenn ein Auto aus einem hell beleuchteten Bildbereich in einen dunklen Schatten hineinfährt), so tritt das Korrespondenzproblem in einer wesentlich verschärften Form auf, weil dann eine Zuordnung der Pixel anhand ihrer Helligkeit bzw. ihres Helligkeitskontrastes nicht mehr möglich ist.

Einen Ansatz zur Minimierung des Korrespondenzproblems, der kein Objektwissen verlangt, stellt die Erweiterung des bilokalen Reichardt-Detektors auf einen multilokalen Detektor mit größerem räumlichem und zeitlichem Integrationsbereich dar, wie sie z.B. von Fenske et al. [1995] vorgeschlagen wurde. Da in diesem Modell mehrere Eingänge in einer faktischen UND-Verknüpfung ausgewertet werden, ist hier die Signifikanz einer Detektorantwort deutlich größer als beim bilokalen Detektor. Ein Nachteil dieses Ansatzes liegt in der notwendigen Bereitstellung relativ langer Verzögerungszeiten (Fenske et al. vergleichen bis zu 7 Kameraframes, also 280 ms).

Für diese Arbeit wurde das – vom Ansatz her vergleichbare – Modell von Schott [1999] übernommen. Hier werden mehrere, als bilokale Bewegungsdetektoren geschaltete Marburger Modellneurone, über modulatorische *Linking*-Verbindungen miteinander gekoppelt. Rechtwinklig zur Vorzugsrichtung des Detektors erfolgt die Kopplung unverzögert; dadurch wird das Gestaltgesetz des *gemeinsamen Schicksals* unterstützt. In der geschätzten Bewegungsrichtung hingegen erhält die Kopplung eine Verzögerung τ, die genau der Zeit entspricht, die das Objekt bei der Vorzugsgeschwindigkeit des Detektors bräuchte, um die Distanz zum nächsten in Bewegungsrichtung liegenden Detektor zurückzulegen (delayed linking). Durch diese Kopplung erhält der folgende Detektor genau die Vorinformation (in Form einer *Voraktivierung*), die bei Zutreffen der Trägheitsannahme mit der tatsächlich zu erwartenden Aktivierung übereinstimmt. Er kann so schneller reagieren; außerdem können durch Rauschen oder ungleichmäßige Beleuchtung verursachte Schwankungen in der Aktivierung bis zu einem gewissen Grad ausgeglichen werden. Die in Kap. 6 zur Kopplung von Kantendetektoren angestellten Überlegungen gelten hier sinngemäß.

Abb. 3.19 zeigt das Transientensystem im Überblick.

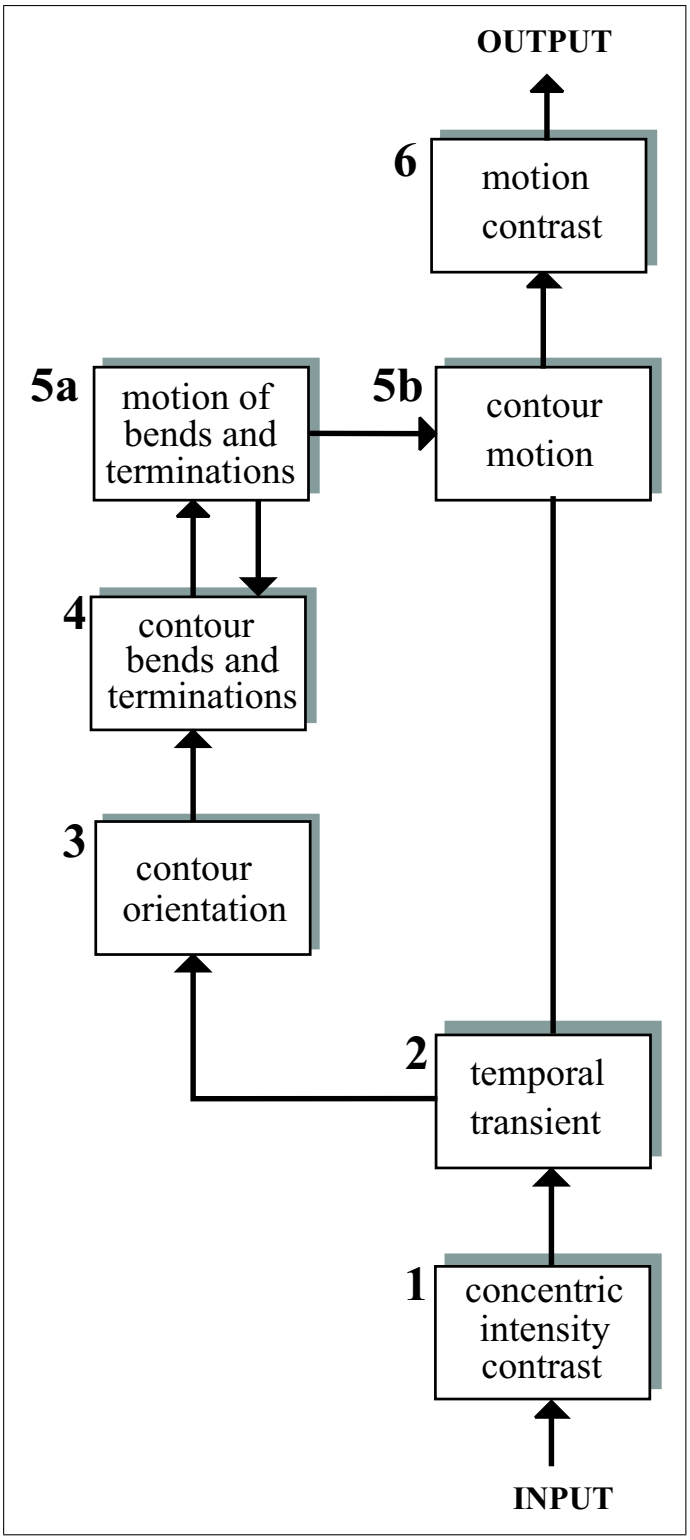


Abbildung 3.19: Aufbau des Transientensystems. Die von den Y-Zellen der Vorverarbeitung detektierten lokalen Transienten werden von den Bewegungsdetektoren getrennt für den ON- und OFF-Pfad ausgewertet. Diese haben ein überlappendes Geschwindigkeitstuning in drei Auflösungsstufen, so daß insgesamt ein breiter Geschwindigkeitsbereich abgedeckt werden kann. Für jede der sechs Hauptrichtungen des pseudohexagonalen Gitters und jede Auflösungsstufe ist eine vollständige retinotope Karte von Bewegungsdetektoren vorhanden. Die so ermittelten Geschwindigkeitsfelder werden nochmals auf lokale Kontraste hin untersucht: Das Ergebnis ist in den *Geschwindigkeitskontrast*-Schichten repräsentiert. (Aus: [Schott, 1999])

3.5 Die Aufmerksamkeitssteuerung

3.5.1 Anforderungen an die Aufmerksamkeits- und Blicksteuerung

Für eine effiziente Segmentierung komplexer dynamischer Szenen ist eine der Segmentierung vorgeschaltete Informationsreduktion unabdingbar. Diese wird zum größten Teil von der in Kap. 3.2 beschriebenen Vorverarbeitung geleistet, die den gesamten Strom von Eingangsbildern in einem System retinotoper Neuronenschichten bearbeitet. Eine weitere Informationsreduktion läßt sich durch eine gezielte Auswahl des relevanten Bildausschnitts erreichen. Diese Funktion wird, wie beim Menschen, durch eine Blicksteuerung ermöglicht, die ihren Input von einem Aufmerksamkeitssystem erhält, dessen Eigenschaften bei der Auswahl des Aufmerksamkeitsfokus auf die nachfolgende Aufgabe zugeschnitten sind (vgl. dazu auch Kap. 2.5.2). Die Aufgabe der Bildsegmentierung und -analyse vereinfacht sich durch die Begrenzung auf einen sinnvoll vorgewählten Ausschnitt in mehrfacher Hinsicht:

1. In jedem Kameraframe müssen weniger Pixel bearbeitet werden, d.h. die retinotopen Schichten des Netzwerks benötigen weniger Neurone zur Abdeckung des Eingangsbildraums.

2. Reale Szenen beinhalten i.a. eine große Anzahl von Objekten; läßt sich ein relevanter Ausschnitt angeben, so reduziert sich diese Anzahl in den meisten Fällen erheblich. Wie sich in Kap. 4.1 zeigen wird, wird dadurch die Segmentierungsdynamik ebenfalls robuster.

3. Durch das Zentrieren von Objekten auf der Retina wird weitgehende *Translationsinvarianz* erzeugt. Ein Mustervergleich zur Objekterkennung ist viel leichter durchzuführen, wenn die ungefähre Lage des Objekts im Bild mit der in der Vorlage übereinstimmt.

Der erste Punkt ist auch im Hinblick auf eine Hardware-Implementation von großer Bedeutung; aufgrund der begrenzten Kommunikationsbandbreite ist keine neuronale Hardware unbegrenzt skalierbar, auch wenn sie vom Design her für den Parallelbetrieb mehrerer Einheiten ausgelegt ist. Selbstverständlich ist das Problem mit der Einführung einer Blicksteuerung nicht gelöst, sondern nur in zwei Teilprobleme zerlegt worden: Das ursprüngliche Segmentierungsproblem besteht in vereinfachter Form fort; hinzugekommen ist die Aufgabe, aus relativ wenigen, voranalysierten Eingangsdaten den relevanten Bildbereich zu identifizieren. Dies leistet die Aufmerksamkeits- und Blicksteuerung, und zwar sowohl bei einer vollständigen Neuorientierung als auch während laufender Veränderungen einer Szene.

Im Organismus ist das Prinzip der selektiv hochaufgelösten Verarbeitung u.a. durch die weitgehende Beschränkung der Bildanalyse auf die Fovea in Kombination mit beweglichen Augen verwirklicht; gleichzeitig läßt sich damit die genannte Translationsinvarianz erreichen. Beim Menschen wird dabei die Blickrichtung durch ein komplexes Wechselspiel von datengetriebenen *Low-Level*-Einflüssen und kortikalen Mechanismen gesteuert: Während des natürlichen Sehens löst z.B. ein transienter visueller Reiz (unabhängig ob

stehend oder bewegt) i.a. sofort eine Sakkade zum Reizort aus. Dieses Verhalten kann zwar teilweise durch kortikalen Einfluß verändert werden, jedoch ist dafür normalerweise eine bewußte Willensanstrengung erforderlich. Überhaupt nicht durch bewußten Einfluß zu verändern ist dagegen das in Kap. 2.5.2 beschriebene Phänomen der veränderten zeitlichen Verarbeitungseigenschaften der Wahrnehmung unter Aufmerksamkeit.

Diese Feststellungen betonen die große Rolle, die auch beim Menschen die unbewußten Prozesse für die Ausrichtung der Aufmerksamkeit spielen. Das Primat der datengetriebenen Einflüsse für die Aufmerksamkeit ist jedoch auch aus funktioneller Sicht zu verstehen: Erst wenn ein relevantes Objekt auf die Fovea zentriert wird (und bei Bewegung für eine gewisse Zeit dort gehalten wird), kann eine genauere formbezogene Analyse ablaufen.[1] Aus diesen Überlegungen heraus lassen sich für die Aufmerksamkeitssteuerung eines technischen Systems folgende **Anforderungen** formulieren:

1. Aus dem Eingangsdatenstrom muß nach kurzer Vorverarbeitung eindeutig ein Blickziel ausgewählt werden.
2. Diese Entscheidung sollte grundsätzlich robust gegen kleinere Störungen sein, um ein instabiles Verhalten zu vermeiden.
3. Die Blickzielauswahl muß anschließend in eine Sakkade umgesetzt werden, die den ausgewählten Bildbereich auf die Fovea zentriert.
4. Bei bewegten Objekten muß anschließend an die Sakkade eine Verfolgung stattfinden.
5. Nach erfolgter Segmentierung/Erkennung soll ein neues Blickziel ausgewählt werden.
6. Plötzliche Änderungen (starke Transienten) in der Szene sollen ebenfalls eine Neuorientierung des Systems auslösen, um eine schnelle Reaktion auf Änderungen zu ermöglichen.
7. Es sollte eine Schnittstelle für die spätere Einbeziehung höherer Verarbeitungsebenen vorhanden sein (*Top-Down*-Wechselwirkung). Insbesondere sind hier intentionale und wissensbasierte (modellgetriebene) Steuerungsmöglichkeiten zu berücksichtigen.
8. Die Integration weiterer sensorischer Modalitäten, wie beispielsweise der akustischen oder somatosensorischen, sollte möglich sein.

3.5.2 Die Modelle von Amari und Kopecz

Ein Modell, das die oben genannten Anforderungen erfüllt und zudem relativ einfach zu implementieren ist, wurde von Amari [1977] vorgeschlagen und in abgewandelter Form von Kopecz [1995] zur quantitativen Vorhersage sakkadischer Reaktionszeiten verwendet.

[1] Hier wird die Zirkularität des Problems der Szenensegmentierung deutlich: Auch ein großer Vorrat an bekannten Objekten in Verbindung mit einem leistungsfähigen Assoziativspeicher ist weitgehend nutzlos, wenn eine Segmentierung der (komplexen) Szene unmöglich ist.

Beide Varianten bestehen im wesentlichen aus einem einschichtigen *neuronalen Feld* mit lateraler, räumlich homogener Wechselwirkung.[2] In der Terminologie der neuronalen Felder bekommt jeder Ort $\vec{x}$ eine zeitabhängige interne Aktivierung $u(\vec{x},t)$ zugeordnet. Diese dient als Argument einer Sigmoid- oder Schwellenfunktion und erzeugt so das Ausgangssignal am jeweiligen Ort. Aktiviert werden die Neurone einerseits von äußeren Einflüssen und andererseits vom Input aus ihrer lateralen Wechselwirkung. Schließlich bestimmt die Zeitkonstante τ_{att} die Reaktionszeit des Systems auf äußere Reize. Die gesamte Dynamik des Systems läßt sich dann durch eine einzige Integro-Differentialgleichung 1. Ordnung beschreiben:

$$\tau_{att}\frac{\partial u(\vec{x},t)}{\partial t} = -u(\vec{x},t) + \int_{R^2} w(\vec{x}-\vec{x}')S[u(\vec{x}',t)]\,d\vec{x}' + f_{vis}(\vec{x},t) \tag{3.12}$$

wobei S die erwähnte Sigmoid- oder Schwellenfunktion ist und $w(\vec{x}-\vec{x}')$ den abstandsabhängigen Wechselwirkungsterm bezeichnet. Letzterer ist in beiden Modellen so angelegt, daß benachbarte Orte sich gegenseitig erregen, weiter entfernte sich dagegen in ihrer Aktivität hemmen. Der Unterschied liegt darin, daß bei Amari die Reichweite der Wechselwirkung begrenzt ist, während im Kopecz-Modell jeder Ort mit jedem wechselwirkt, entsprechend einem vollverbundenen neuronalen Netz. Dazu verwendet Amari eine Mexikanerhut-Funktion, Kopecz eine Gaußfunktion mit negativem konstantem Offset $-h_{sel}$:

$$w_{Amari}(\vec{x}-\vec{x}') = \frac{1}{\sqrt{2\pi}\,\sigma^3}\left(\frac{|\vec{x}-\vec{x}'|^2}{\sigma^2}-1\right)e^{-\frac{1}{2}\frac{|\vec{x}-\vec{x}'|^2}{\sigma^2}} \tag{3.13}$$

und

$$w_{Kopecz}(\vec{x}-\vec{x}') = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{1}{2}\frac{|\vec{x}-\vec{x}'|^2}{\sigma^2}} - h_{sel} \tag{3.14}$$

Die zugehörige Verschaltungsstruktur ist in Abb. 3.20c dargestellt. Die lokale Exzitation führt dazu, daß eine einmal im Netz vorhandene, lokalisierte Aktivität eine gewisse Tendenz zur Selbsterhaltung hat. Diese kommt durch die gegenseitige Erregung benachbarter Neurone zustande. Ob eine stationäre Selbsterhaltung von Aktivität ohne äußeren Input tatsächlich möglich ist, hängt von den Netzwerk-Parametern ab, insbesondere von der Stärke der Exzitation und der Größe des exzitatorisch angekoppelten Bereichs der Wechselwirkung.

Die Inhibition der weiter entfernten Neurone durch lokale Aktivität führt gleichzeitig zu einer räumlichen Begrenzung der aktiven Zone: dort, wo die Summe der inhibitorischen Einflüsse die Summe der Exzitation überwiegt, liegt die Grenze des sich effektiv selbst erregenden Bereichs. Im Ergebnis bilden sich (bei konstantem Input) stabile Zonen von räumlich begrenzter Aktivität, die ich im folgenden als *Aktivitätblobs* oder kurz *Blobs*

[2]Der Begriff des neuronalen Feldes wird dazu verwendet, räumlich oder topologisch geordnete neuronale Netze mit vielen gleichartigen Elementen in einer Kontinuumsnäherung zu behandeln. Die kontinuierliche Schreibweise ermöglicht oftmals eine weitergehende mathematische Behandlung. In numerischen Simulationen muß ohnehin eine Diskretisierung vorgenommen werden, so daß hier kein Unterschied zwischen einem neuronalen Netz mit vielen Elementen und einem diskretisierten Feld mehr besteht.

bezeichne. Da im Modell von AMARI die Reichweite der Wechselwirkung und damit der Inhibition begrenzt ist, können hier ohne weiteres mehrere solcher Blobs in ausreichendem räumlichen Abstand koexistieren – es läuft dann lokal mehrfach die gleiche Dynamik ab. Im Gegensatz dazu wird dies bei KOPECZ durch die mit unbegrenzter Reichweite arbeitende Inhibition verhindert. Ein einmal etablierter Aktivitätsblob inhibiert alle anderen Neurone in der Schicht.

3.5.3 Erzeugung eines geeigneten Eingangssignals für die Blicksteuerung

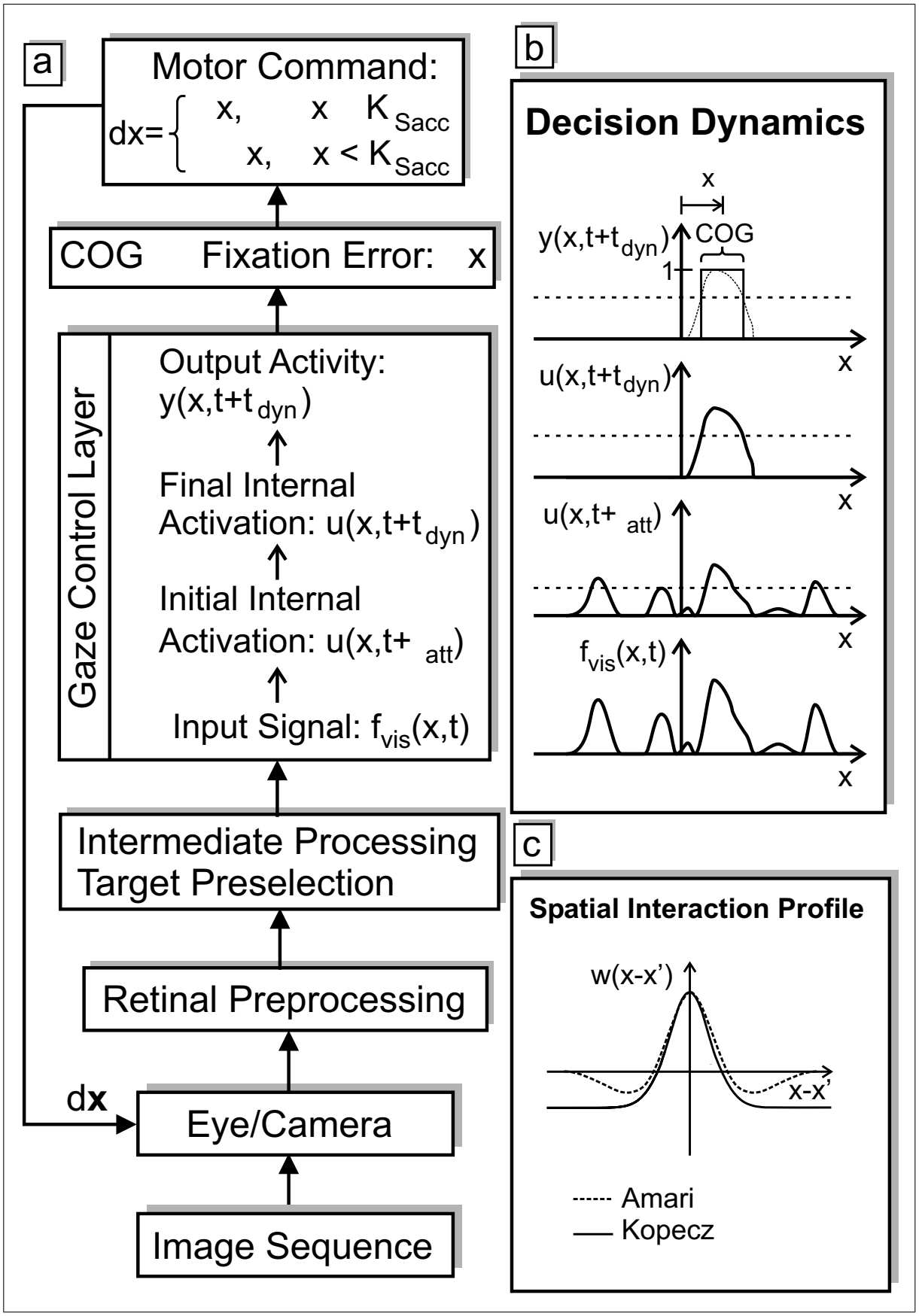
Als visueller Input $f_{vis}(\vec{x}, t)$ wird ein zeitabhängiges 'Aufmerksamkeitsgebirge' (*saliency map*) verwendet, das bei der hier angewandten, datengetriebenen Arbeitsweise von der Vorverarbeitung erzeugt werden muß.[3] Sollen, wie bei KOPECZ, psychophysische Experimente nachgebildet werden, so ist dieser Schritt i.a. unproblematisch, da hier meistens mit isolierten Lichtpunkten als Reiz gearbeitet wird und keine zusätzliche Vorverarbeitung erforderlich ist, um den Ort potentieller Blickziele festzustellen. Ebenso kann als Reizstärke einfach die Helligkeit bzw. der lokale Kontrast der Lichtpunkte herangezogen werden.

Anders ist dies bei der Analyse realer Szenen. Hier besteht eine entscheidende Aufgabe darin, die Vorverarbeitung aus der Vielfalt der angebotenen Reize eine geeignete Vorauswahl über mögliche Blickziele treffen zu lassen. Es ist also ein *Anforderungsprofil* für die Blicksteuerung zu formulieren, das dann mittels einer geeigneten Vorverarbeitung umgesetzt wird.

Für eine allgemein angelegte Analyse bewegter Szenen, die – wie oben beschrieben – möglichst wenige Annahmen über die Beschaffenheit einzelner Objekte treffen soll, kommt eine einfache Analyse aufgrund lokalen Intensitätskontrastes prinzipiell ebenfalls in Fra-

[3] Mit Vorverarbeitung wird hier nicht nur der retinale Teil bezeichnet, sondern alle Verarbeitungsschritte, die dem Finden des Blickziels vorausgehen.

Abbildung 3.20: **(a)** Signalfluß im Aufmerksamkeitssystem. Die vorgeschalteten Verarbeitungsstufen (insbesondere das Transientensystem) generieren einen retinotopen Input, der als Aufmerksamkeitskarte für die Blicksteuerung dient. **(b)** Entscheidung für ein Blickziel durch die Netzwerkdynamik. Die externe Anregung $f_{vis}(\vec{x}, t)$ wird, nach der zeitlichen Tiefpaßfilterung am Eingang (Zeitkonstante τ_{att}), als Membranpotential $u(\vec{x}, t)$ der Aufmerksamkeitsneurone abgebildet. Die gegenseitige Inhibition entfernter Neurone führt in Kombination mit ihrem Schwellenmechanismus zu einem Winner-Take-All-Wettbewerb zwischen den lokalen Inputmaxima, die durch überschwellige Anregung der Neurone Aktivität in der Aufmerksamkeitsschicht generieren können. Nach wenigen Zeitschritten τ_{dyn} ist der Wettbewerb entschieden; ein einzelner Aktivitätsblob markiert mit seinem Schwerpunkt (COG) den Ort des nächsten Blickziels. Das Beispielsignal ist hier aus Gründen der besseren Übersicht in seiner eindimensionalen Form dargestellt; für die vorliegende Arbeit wurde jedoch die (einfache) Verallgemeinerung auf zwei Dimensionen verwendet. **(c)** Laterale Verschaltungsstruktur innerhalb der Aufmerksamkeitsschicht. Jedes aktive Neuron erregt seine unmittelbare Nachbarschaft und hemmt weiter entfernt liegende Neurone. Im Modell von AMARI (gestrichelte Linie) ist die Reichweite der Wechselwirkung begrenzt; die Bildung mehrerer Aktivitätsblobs in ausreichender Entfernung voneinander ist grundsätzlich möglich. Bei der Variante von KOPECZ (durchgezogene Linie) erreicht die Inhibition dagegen immer das gesamte Feld, weshalb höchstens eine Region stabil aktiviert sein kann.


a
Motor Command:
dx= { x, x K Sacc
x, x < K Sacc
COG Fixation Error: x
Gaze Control Layer
Output Activity:
y(x,t+t dyn)
Final Internal
Activation: u(x,t+t dyn)
Initial Internal
Activation: u(x,t+ att)
Input Signal: f vis(x,t)
Intermediate Processing
Target Preselection
Retinal Preprocessing
dx
Eye/Camera
Image Sequence
b
Decision Dynamics
x
y(x,t+t dyn)
COG
1
x
u(x,t+t dyn)
x
u(x,t+ att)
x
f vis(x,t)
x
c
Spatial Interaction Profile
w(x-x')
x-x'
Amari
Kopecz

ge. Der Vorteil liegt in der einfachen Umsetzung; die retinotopen Ausgangssignale der retinalen Vorverarbeitung können direkt als Aufmerksamkeitskarte für die Blicksteuerung dienen. Der große Nachteil einer solch einfachen Analyse ist in der geringen Selektivität der Auswahl zu sehen: In der Beispielszene *Durlacher Tor* werden neben den eigentlich interessierenden Autos auch die kontraststarken Fahrbahnmarkierungen als Blickziele ausgewählt (s. Kap. 5).

Möchte man die Auswahl ausschließlich auf bewegte Objekte konzentrieren, dann bietet sich die Nutzung der vom Transientensystem erzeugten Signale an. Diese haben folgende Vorteile:

1. Jedes *bewegte* Objekt mit ausreichendem Intensitätskontrast erzeugt dort eine Antwort, da jede Bewegung mit zeitlichen Helligkeitsänderungen einhergeht.

2. Stillstehende Bildteile (Hintergrund bei fester Kamera) erzeugen keine Antwort im Transientensystem und können somit auch keine Aktivität in der Aufmerksamkeitsschicht mehr hervorrufen.

3. Selbst bei nicht stillstehendem Hintergrund (z.B. während einer Folgebewegung) liefert das Transientensystem mit dem *Bewegungskontrast* ein retinotopes Signal, das ein Herausfiltern der *unterschiedlich* zum Hintergrund bewegten Bereiche erlaubt.

Die Nachteile bei einer Nutzung des Transientensystems liegen einerseits in der erforderlichen aufwendigen Vorverarbeitung vor der Blicksteuerung, andererseits in der ausschließlichen Beschränkung auf bewegte Bildteile. Der zweite Punkt läßt sich, falls erforderlich, durch eine geeignete Kombination statischer und transienter Signalanteile umgehen. Um den Aufwand (besonders im Hinblick auf einen technischen Einsatz) zu begrenzen, können je nach Problemstellung nur Teile des Transientensystems zum Einsatz kommen.

Grundsätzlich kann jedes lokalisierte Signal in die retinotope Aufmerksamkeitsschicht eingespeist werden. Dabei läßt sich bei Überlagerung mehrerer Informationskanäle durch geeignete Gewichtung fast jedes gewünschte Verhalten herstellen. Diese Eigenschaft der *Allgemeinheit* ist in dreierlei Hinsicht besonders interessant:

1. **Technischer Bereich:** Bestimmte Muster (die nicht notwendig durch neuronale Verfahren gefunden werden müssen), können einen eigenen 'Aufmerksamkeitskanal' erhalten, d.h. es kann z.B. durch Mustervergleich (*'Template Matching'*) gezielt nach bestimmten Werkstücken gesucht und die Kamera dann auf diese ausgerichtet werden.

2. **Sensorische Integration:** Es ist ohne weiteres möglich, Informationen aus anderen sensorischen Modalitäten zu berücksichtigen, z.B. der akustischen. Dabei läßt sich durch eine geeignete Gewichtsverteilung am Eingang der Aufmerksamkeitsschicht auch eine unterschiedliche räumliche Trennschärfe der verschiedenen Kanäle berücksichtigen.

3. **Wechselwirkung mit höheren Ebenen:** Höhere Verarbeitungsebenen können einen eigenen, direkten Zugang zur Blicksteuerung erhalten. Durch direkte oder modulatorische (Vor-)Aktivierung können dann Aktionen *top-down* anstatt datengetrieben ausgelöst werden, entsprechend kortikalen Einflüssen im natürlichen System.

Der letzte Punkt ist im Modell von KOPECZ [1995] bereits enthalten: Die *Absicht* der Versuchsperson, als Folge der Instruktion den Fixpunkt zu fixieren, wurde als zusätzlicher additiver Term im Input berücksichtigt, um die gemessenen Reaktionszeiten erklären zu können.

3.5.4 Umsetzung der Aufmerksamkeitssteuerung mit Marburger Modellneuronen

Für die Verwendung in der vorliegenden Arbeit kam von den beiden o.a. Modellvarianten nur diejenige von KOPECZ in Frage, da nur eine Kamera auszurichten und damit immer ein *eindeutiges* Blickziel auszuwählen ist.

Um dieses Modell mit Marburger Modellneuronen in einer zweidimensionalen Schicht umzusetzen, wurde daher zunächst die laterale Verschaltung auf zwei Dimensionen verallgemeinert. An die Stelle der allgemeinen nichtlinearen Funktion S aus Gl. 3.12 tritt beim Marburger Modellneuron als Spezialfall einer solchen Nichtlinearität die zeitabhängige Feuerschwelle $\Theta(t)$. Die Rolle der Zeitkonstante τ_{att} übernimmt die Feeding-Zeitkonstante τ_F des Marburger Modellneurons.

Die qualitativen Eigenschaften eines solchen Netzwerkes unterscheiden sich kaum von den bei KOPECZ beschriebenen: Es wird immer ein eindeutiges Blickziel ausgewählt; diese Auswahl ist robust gegen kleine Störungen und zeigt eine gewisse (einstellbare) Hysereseeigenschaft, vgl. Kap. 3.5.9.

3.5.5 Steuerung der Blickbewegungen

3.5.5.1 Räumliche Integration der Aktivität in der Aufmerksamkeitsschicht

Die Herausbildung eines einzelnen Aktivitätsblobs ist im allgemeinen innerhalb weniger Simulationschritte abgeschlossen ($\tau_{dyn} < 20\,ms$), läuft also auf einer deutlich schnelleren Zeitskala als die zeitliche Abtastung der Eingangsbildfolge. Deshalb kann im Prinzip ständig die momentane Aktivität in der Aufmerksamkeitsschicht zur Bestimmung eines Blickziels benutzt werden; eine gesonderte Überwachung des 'Entscheidungszeitpunkts' ist nicht erforderlich. Dazu ist zunächst eine *räumliche Integration* über die Aktivität in der Aufmerksamkeitsschicht erforderlich. Da die stabile Aktivitätsregion räumlich immer zusammenhängt, genügt es für die *Lokalisierung* des Aufmerksamkeitsfokus, den Schwerpunkt des Aktivitätsblobs zu bestimmen. Der Abstand zum retinalen Koordinatenursprung wird als *angezeigter Fixationsfehler* $\tilde{x}^{Err}(t)$ zum jeweiligen Zeitpunkt t bezeichnet. Die Verwendung des Schwerpunktes der Aktivität als Blickziel ist einerseits rechnerisch einfach zu realisieren, andererseits steht sie auch in Übereinstimmung mit neueren Ergebnissen zu Blickzielen von Sakkaden bei komplexen Reizkonfigurationen [MCGOWAN

ET AL., 1998]. Die *angezeigte Targetposition* $\widetilde{\vec{x}}^T(t)$ berechnet sich also vorbehaltlich einer zeitlichen Integration wie folgt:

$$\widetilde{\vec{x}}^T(t) = \int_{R^2} \vec{x} S[u(\vec{x}, t)]\, d\vec{x} \tag{3.15}$$

und daraus der momentane angezeigte Fixationsfehler $\widetilde{\vec{x}}^{Err}(t)$

$$\widetilde{\vec{x}}^{Err}(t) = \widetilde{\vec{x}}^T(t) - \vec{x}^F(t) \tag{3.16}$$

wobei $\vec{x}^F(t)$ die Position des Kamerafixpunkts (also die Blickrichtung) zur Zeit t angibt. Die wahre Position $\vec{x}^T(t)$ eines Blickziels (etwa bei einer Verfolgungsaufgabe) muß keineswegs mit der von der Aufmerksamkeitsschicht angezeigten übereinstimmen; zur Unterscheidung wird daher die Bezeichnung $\widetilde{\vec{x}}^T(t)$ für letztere verwendet (und analog für $\vec{x}^{Err}(t)$ bzw. $\widetilde{\vec{x}}^{Err}(t)$).

3.5.5.2 Zeitliche Integration und Abtastung des Aufmerksamkeitssignals

Verfolgt ein Mensch ein gleichmäßig bewegtes Objekt mit den Augen, so führen diese dabei ebenfalls eine stetige Bewegung aus. Derartige Folgebewegungen können mit einer realen Kamera gut nachgebildet werden; in der numerischen Simulation auf Pixelbildern beträgt ohne zusätzliche Interpolation die kleinste mögliche Bewegung 1 Pixel. Für eine möglichst gute Annäherung an die stetige Verfolgung ist deshalb zu fordern, daß der Zeittakt der Kamerasignale so ausgelegt ist, daß während einer glatten Folgebewegung möglichst keine größeren Kameraschritte als 1 Pixel auftreten sollten.

Diese Forderung läßt sich durch einen ausreichend schnellen Zeittakt bei der Erzeugung der Kamerasignale leicht erfüllen, zumal die kürzeste im System verwendete Taktzeit mit 1 *bin* eine um etwa zwei Größenordungen schnellere Zeitskala als die äußeren Vorgänge repräsentiert. Die naheliegende Lösung, Kamerasignale im Simulationstakt zu erzeugen, birgt allerdings ein anderes Problem: Bei der Verwendung impulscodierender Neurone zur Realisierung der Aufmerksamkeitssteuerung entsteht die Ausgangsaktivität auch im stationären Zustand durch die Überlagerung vieler einzelner Spike-Ereignisse. Auch bei starker Anregung im Blob und somit hoher Feuerrate der Neurone unterliegt die Aktivitätsverteilung einer schnellen Fluktuation. Würde man die Aktivität der Aufmerksamkeitsschicht also im Abstand von 1 *bin* abtasten, so erhielte man (nach der Schwerpunktberechnung) ein Positionssignal, das ständig um den eigentlichen Aktivitätsschwerpunkt schwankt. Um diese schnellen Fluktuationen zu eliminieren, genügt eine zeitliche Integration über einige Simulationszeitschritte, entsprechend der Feuerrate der aktiven Neurone in der Aufmerksamkeitsschicht. Als Kompromiß, der sowohl schnelle Verfolgung als auch ausreichende Integration (und damit Glättung) ermöglicht, wurde in allen Simulationen ein Integrationsintervall von $\Delta t_{Move} = 8\, bin = 10\, ms$ verwendet. Diese Wahl hat zusätzlich den Vorteil, daß aufgrund des ganzzahligen Verhältnisses zum Zeittakt der Kameraframes eine Synchronisation der Teilsysteme auf einfache Weise erfolgen kann.

Daraus ergibt sich eine um den Faktor 4 gröber abgetastete Zeitachse für die Kamerabewegung als für die Simulation selbst. Die für die Erzeugung der Kamerabewegung

verwendete angezeigte Targetposition lautet mit dieser Integration (in kontinuierlicher Schreibweise):

$$\tilde{\vec{x}}^T(t) = \int_{t-\Delta t_{Move}}^{t} \int_{R^2} \vec{x} S[u(\vec{x},t)]\, d\vec{x}\, d\tau \tag{3.17}$$

Auf eine unterschiedliche Gewichtung der Abschnitte innerhalb des Zeitintervalls Δt_{Move} wurde verzichtet, da diese Zeitskala ohnehin deutlich schneller als die äußeren Abläufe ist und somit eine interne Gewichtung keinen Gewinn mehr bringt.

Für den Übergang zwischen zeitdiskreter und zeitkontinuierlicher Darstellung notieren wir zunächst die diskretisierten Größen im Überblick:

$$\begin{aligned} t_n &= n \cdot \Delta t_{Move} \\ \vec{x}_n^i &= \vec{x}^i(t_n) \\ \Delta\vec{x}_n^i &= \vec{x}_n^i - \vec{x}_{n-1}^i \\ v_n^i &= \dot{\vec{x}}^i(t_n) \approx \frac{\Delta\vec{x}_n^i}{\Delta t_{Move}}, \qquad i \in \{F, T, Err, \tilde{}\} \end{aligned} \tag{3.18}$$

wobei v_n^i die mittlere Geschwindigkeit der Blickbewegung im Intervall $[t_{n-1}, t_n]$ bezeichnet.

3.5.6 Auslösung und Steuerung der Blickbewegungen

Wie bereits erwähnt, ist aufgrund des typischerweise schnellen Ablaufs der Aufmerksamkeitsdynamik ($\tau_{dyn} < \Delta t_{Move}$) im Vergleich zur Zeitskala der Blickbewegungen eine gesonderte Entscheidungsstufe für die Auslösung einer Blickbewegung nicht erforderlich. Zu jedem Zeitpunkt t_n wird lediglich der Schwerpunkt $\tilde{\vec{x}}^T(t_n)$ der Aktivität in der Aufmerksamkeitsschicht ausgewertet und nach Gl. 3.16 der angezeigte Fixationsfehler $\tilde{\vec{x}}_n^{Err}$ berechnet. Aus diesem ist nun die eigentliche Blickbewegung $\Delta\vec{x}_n^F$ zu bestimmen. Aus Gründen, die im folgenden noch klar werden, ist es sinnvoll, hier eine Unterscheidung zwischen einem großen Fixationsfehler, der ein peripher gelegenes Blickziel anzeigt, und einer kleinen Abweichung zu treffen: Im ersten Fall ist eine Sakkade direkt zum Ziel erforderlich, um möglichst sofort eine fehlerfreie Fixation zu erreichen. Der zweite Fall repräsentiert dagegen eine kleine Abweichung, wie sie z.B. während einer Objektverfolgung oder nach einer ungenauen Sakkade auftreten kann. Als ungefähres *a priori*-Kriterium zur Unterscheidung läßt sich der typische Durchmesser des Aktivitätsblobs in der Fixationsschicht heranziehen: Bei einem Fixationsfehler, der kleiner als der Blobdurchmesser ist, kann die Aktivität allein durch 'Wandern' des Blobs und ohne Kompetition zum neuen Inputmaximum gezogen werden. Wettbewerb tritt erst auf, wenn das neue Blickziel so weit vom bisherigen Fixationsort entfernt ist, daß der inhibitorische Anteil der Wechselwirkung greift.

Das zur Unterscheidung der beiden Typen von Blickbewegungen verwendete Kriterium wird als K_{Sacc} bezeichnet. Ist der Fixationsfehler größer, so wird einfach eine Sakkade zum neuen Blickziel generiert. Ist er kleiner, so wird ebenfalls eine Bewegung zum Blickziel hin ausgelöst, die allerdings nur einen Bruchteil $\hat{\alpha}$ des angezeigten Fixationsfehlers $\tilde{\vec{x}}^{Err}$

beträgt:

$$\Delta\vec{x}_n^F = \begin{cases} \tilde{\vec{x}}_n^{Err}, & \tilde{\vec{x}}_n^{Err} \leq K_{Sacc} \\ \hat{\alpha} \cdot \tilde{\vec{x}}_n^{Err}, & \tilde{\vec{x}}_n^{Err} > K_{Sacc} \end{cases} \quad \text{mit} \quad \hat{\alpha} \leq 1 \tag{3.19}$$

Die Bedeutung des Verfolgungsparameters $\hat{\alpha}$ wird in Kap. 3.5.8 genauer behandelt. Die beiden folgenden Abschnitte erläutern die Konsequenzen, die sich aus dem Dauerbetrieb der Blicksteuerung in einer geschlossenen Schleife für die Behandlung von Sakkaden sowie Folgebewegungen ergeben.

3.5.7 Sakkaden

Wird zum Zeitpunkt t_n eine Sakkade $\Delta\vec{x}_n^F$ ausgeführt, dann führt im retinotopen Koordinatensystem das gesamte Eingangsbild eine entgegengesetze Bewegung aus. Die Aktivität in der Aufmerksamkeitsschicht besteht jedoch aufgrund der Selbsterhaltung zunächst am gleichen Ort fort. Unmittelbar nach einer Sakkade ist also der angezeigte Fixationsfehler nicht Null oder nahe Null, sondern aufgrund der veränderten Koordinatentransformation zwischen Retina- und Weltkoordinaten ungefähr gleich demjenigen Fehler, der zur Sakkade geführt hat: $\Delta\vec{x}_n^F \approx \Delta\vec{x}_{n+1}^F$. Abb. 3.21 veranschaulicht diesen Sachverhalt.

Da auch im natürlichen System die Neuronenschichten des *Superior Collicullus* retinotop organisiert sind, muß dieser ebenfalls von diesem Effekt betroffen sein. Experimentelle Ergebnisse von Munoz und Wurtz [1993a] deuten darauf hin, daß bei Katzen während der Sakkade eine Aktivitätswelle von der peripheren Abbildung des Blickziels zum 'Fixpunkt' der Blicksteuerungsschicht läuft und so die Aktivität nach der Sakkade wieder in der Mitte des Gesichtsfeldes ihr Maximum hat.

3.5.7.1 Sakkadische Suppression

Ein solches selbständiges Zurücklaufen der Aktivität zum Zentrum der retinotopen Blicksteuerungsschicht läßt sich im Prinzip durch eine geeignete homogene, aber anisotrope laterale Verschaltung erreichen [Schierwagen, 1996]. Diese verändert aber gleichzeitig die Eigenschaften der Kompetitionsdynamik, so daß die Detektion eines peripheren Targets nicht mehr ungestört ablaufen kann – der Blob wird ständig zur Mitte gezogen. Für die vorliegende Arbeit wurde daher ein anderer Weg zur Behebung des postsakkadischen Fixationsfehlers gewählt.

An Stelle einer Verschiebung zum Zentrum wird der Output der retinalen Vorverarbeitung für eine gewisse Zeit T_{Sup} unterdrückt, d.h. es findet eine *sakkadische Suppression* statt. Sinnvolle Werte für T_{Sup} liegen im Bereich 30–50 ms, also in einem Bereich, der die Echtzeitfähigkeit des Systems im allgemeinen nicht gefährdet. Während dieser Zeitspanne kann die Aktivität in der Aufmerksamkeitsschicht relaxieren; anschließend wird sie neu aufgebaut. Da bei größeren Verschiebungen der Blickrichtung ohnehin Neurone anderen Bildorten zugeordnet werden und somit eine erneute Segmentierung der Szene erforderlich ist, bietet sich hier eine elegante Möglichkeit, die in Kap. 4.1 dargestellten aufmerksamkeitsabhängigen Latenzen in den Verarbeitungsablauf des Systems einzuführen (s. Kap. 4.4).

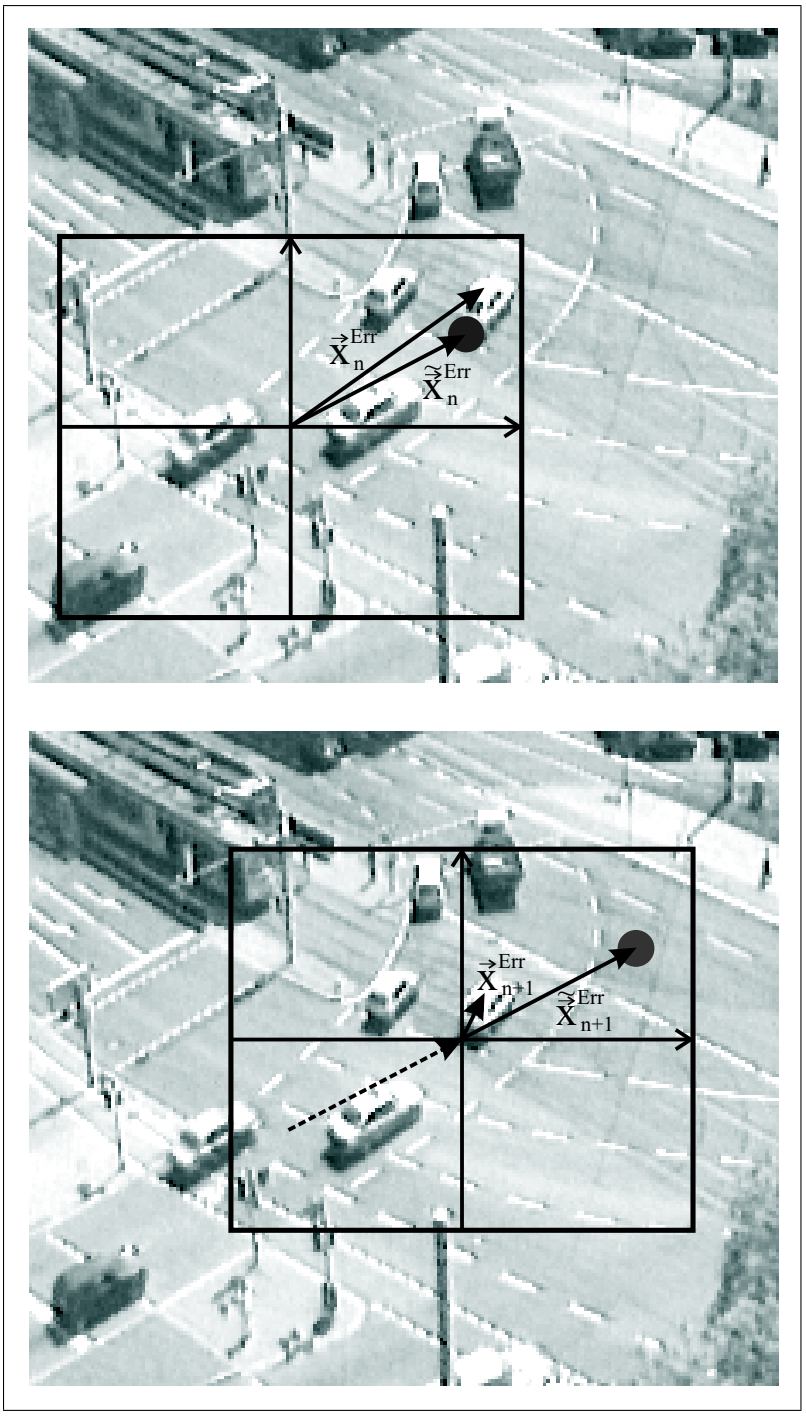


Abbildung 3.21: Für die Blicksteuerung relevante Größen in einer realen Szene. Der Kasten markiert das Gesichtsfeld der Kamera; die Achsen spannen das retinotope Koordinatensystem auf. Im Beispiel erfaßt die Aufmerksamkeitsdynamik das Blickziel (Auto) nicht exakt, so daß tatsächlicher und angezeigter Fehler leicht voneinander abweichen (oberes Bild). Da der angezeigte Fehler das Kriterium K_{Sacc} zur Auslösung einer Sakkade deutlich übersteigt, wird eine Blickbewegung zur angezeigten Position hin ausgeführt. Wie im eindimensionalen Beispiel aus Fig. 3.22 ist die aus der Aktivität der Aufmerksamkeitsschicht gewonnene Fehlerangabe unmittelbar nach der Sakkade unsinnig. Erst nach einer gewissen Zeitspanne T_{Sup}, in der der Input unterdrückt wird, kann sich die Aktivitätsverteilung entsprechend den neuen Gegebenheiten aufbauen.

3.5.8 Glatte Folgebewegungen

Bei glatten Folgebewegungen, die kontinuierlich bewegte Objekte im Blickzentrum halten sollen, tritt der gleiche Effekt auf, muß aber aus funktionellen Gründen anders behandelt werden. Prinzipiell besteht in der zeitdiskreten Darstellung auch eine glatte Folgebewegung aus vielen Einzelbewegungen, so daß das oben gesagte auch für diesen Fall gilt. Im Gegensatz zu einer größeren Sakkade soll hier aber gerade keine Neusegmentierung der Szene erfolgen, sondern eine eventuell bestehende Segmentierung möglichst erhalten bleiben (vgl. auch Kap. 4.1).

Aus diesen Gründen verbietet sich eine Suppression des Inputs für Folgebewegungen. Statt dessen wird die Blickrichtung entsprechend Gl. 3.19 in jedem Zeitschritt T_n nur um einen Bruchteil $\hat{\alpha}$ des angezeigten Fixationsfehlers nachgeführt. Wie man sich leicht anschaulich klarmachen kann, kann dieses Vorgehen bei einem gleichmäßig bewegten Target sogar dazu führen, daß der Aktivitätsblob kontinuierlich mit dem Target mitwandert, wobei die Blickrichtung mit gleicher Geschwindigkeit nachgezogen wird. Geringe Abweichungen in der Größenordung des Blobdurchmessers werden dabei von der Auswahldynamik ausgeglichen, d.h. die Blob-Position $\widetilde{\vec{x}}^T$ wird ständig an die reale Targetposition $\vec{x}^T$ angepaßt. Im folgenden wird eine quantitative Behandlung dieses Typs von Folgebewegung gegeben.

3.5.8.1 Mathematische Behandlung der Folgebewegung

Abb. 3.22 veranschaulicht die Größen, die für die mathematische Formulierung der Folgebewegungen benötigt werden. In jedem Abtastintervall, zu dessen Beginn $\widetilde{\vec{x}}^{Err} < K_{Sacc}$ gilt, findet folgender Ablauf statt (alle Größen sind in Weltkoordinaten angegeben):

1. Die Blickrichtung wird um $\Delta\vec{x}^F = \hat{\alpha} \cdot \widetilde{\vec{x}}^{Err}$ nachgeführt.
2. Durch die Nachführung wird der Blob mitbewegt: $\Delta\widetilde{\vec{x}}^T_F = \Delta\vec{x}^F$
3. Das Target bewegt sich um $\Delta\vec{x}^T = v^T \cdot \Delta t_{Move}$ weiter.
4. Die Blob-Position verlagert sich entsprechend dem veränderten Input. Diese Verlagerung $\Delta\widetilde{\vec{x}}_{dyn}$ hängt aufgrund der homogenen Struktur der Aufmerksamkeitsschicht nur vom Abstand zwischen Blob- und Targetposition ab: $\Delta\widetilde{\vec{x}}_{dyn} = \hat{f}_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T)$, wobei die Funktion f_{dyn} das Verhalten der Auswahldynamik beschreibt.

Die Bewegung des Blobs setzt sich also einerseits aus der Nachführung der Blickrichtung $\Delta\vec{x}$ und andererseits aus der Anpassung der Dynamik an den sich verändernden Input zusammen. Der Übergang auf die zeitkontinuierliche Darstellung erfolgt mittels Division der $\Delta\vec{x}^i_n$ durch Δt_{Move}. Es ergibt sich das folgende Differentialgleichungssystem:

$$\begin{aligned} \dot{\widetilde{\vec{x}}}^T &= \dot{\vec{x}}^F + f_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T) \\ \dot{\vec{x}}^F &= \alpha \cdot (\widetilde{\vec{x}} - \vec{x}^F) \end{aligned} \tag{3.20}$$

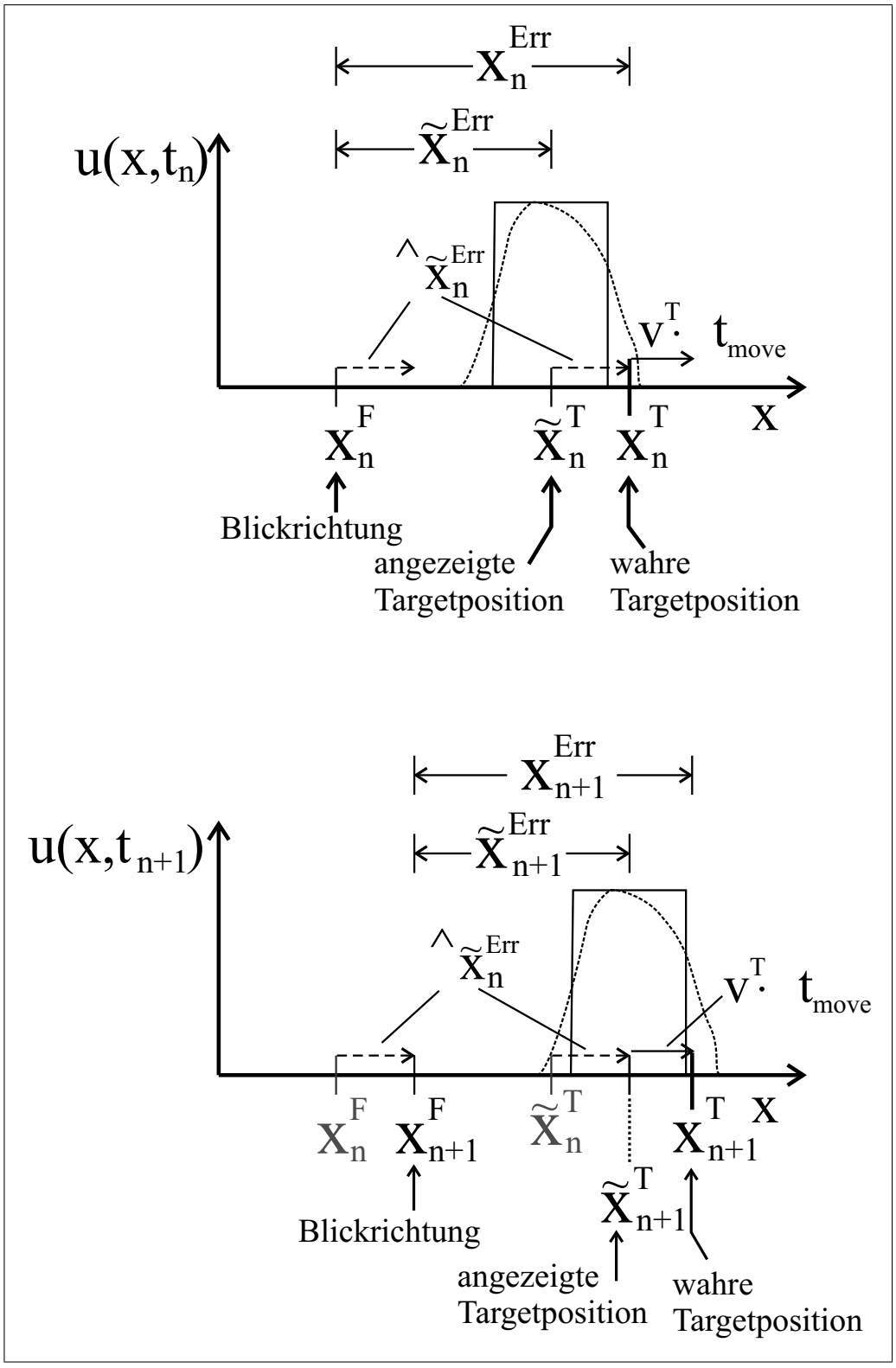


Abbildung 3.22: Darstellung der Größen aus Gl. 3.20 am eindimensionalen Beispiel (in Weltkoordinaten). Während eines Zeitintervalls Δt_{Move} bewegt sich das Target um $v^T \cdot \Delta t_{Move}$ weiter; gleichzeitig wird die Blickrichtung x^F gemäß Gl. 3.19 nachgeführt. Dabei wird die retinotope Aufmerksamkeitsschicht – und damit auch der Aktivitätsblob – um denselben Betrag über die Szene hinweg verschoben. Zu dieser Verschiebung durch die Nachführung kommt dann noch die eigentliche Wanderung des Blobs aufgrund der Aufmerksamkeitsdynamik. Aus der Zusammenfassung aller Beiträge ergibt sich Gl. 3.20.

mit

$$\alpha = \frac{\hat{\alpha}}{\Delta t_{Move}} \qquad \text{und} \qquad f_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T) = \frac{\hat{f}_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T)}{\Delta t_{Move}} \tag{3.21}$$

Dieses Gleichungssystem ist bis auf die allgemein angesetzte Funktion f_{dyn} linear. Während einer Folgebewegung ist allerdings $\vec{x}^T - \widetilde{\vec{x}}^T$ durch die Bedingung $\vec{x}^T - \widetilde{\vec{x}}^T < K_{Sacc}$ nach oben beschränkt (sonst wird eine Sakkade ausgelöst). Deshalb läßt sich für eine lineare Näherung f_{dyn} nach $\vec{x}^T - \widetilde{\vec{x}}^T$ entwickeln; die Terme quadratischer und höherer Ordnung werden vernachlässigt:

$$f_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T) = 0 + \Phi \cdot (\vec{x}^T - \widetilde{\vec{x}}^T) + \ldots \quad \text{mit} \quad \Phi = \left.\frac{\partial f_{dyn}(\vec{x}^T - \widetilde{\vec{x}}^T)}{\partial(\vec{x}^T - \widetilde{\vec{x}}^T)}\right|_{\vec{x}^T = \widetilde{\vec{x}}^T} \tag{3.22}$$

In formaler Matrixschreibweise lautet Gl. 3.20 dann:

$$\begin{pmatrix} \Phi & 0 \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} \widetilde{\vec{x}}^T(t) \\ \vec{x}^F(t) \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{\widetilde{\vec{x}}}^T(t) \\ \dot{\vec{x}}^F(t) \end{pmatrix} = \begin{pmatrix} \vec{x}^T(t) \\ 0 \end{pmatrix} \tag{3.23}$$

3.5.8.2 Homogene Lösungen

Um die allgemeine Lösung dieses inhomogenen Gleichungssystems zu finden, bestimmen wir zunächst die allgemeine Lösung des entsprechenden homogenen Gleichungssystems. Der Ansatz

$$\begin{pmatrix} \widetilde{\vec{x}}^T(t) \\ \vec{x}^F(t) \end{pmatrix} = \begin{pmatrix} \widetilde{C}^T(t) \\ C^F(t) \end{pmatrix} \cdot e^{\lambda t} \tag{3.24}$$

führt auf das Eigenwertproblem

$$\begin{pmatrix} \Phi + \lambda & -\lambda \\ -\alpha & 1 + \lambda \end{pmatrix} \begin{pmatrix} \widetilde{C}^T(t) \\ C^F(t) \end{pmatrix} \cdot e^{\lambda t} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.25}$$

Durch Diagonalisieren ergeben sich die Eigenwerte

$$\lambda_{1;2} = \frac{\alpha - \Phi - 1}{2} \pm \frac{1}{2}\sqrt{(1 + \Phi - \alpha)^2 - 4\Phi} \tag{3.26}$$

Zur übersichtlicheren Notation bezeichnen wir den Radikanden in Gl. 3.26 mit R und zerlegen ihn in Linearfaktoren:

$$\begin{aligned} R &= (\alpha - \Phi - 1)^2 - 4\Phi \\ &= (\alpha - (\Phi + 1 - 2\sqrt{\Phi})) \cdot (\alpha - (\Phi + 1 + 2\sqrt{\Phi})) \end{aligned} \tag{3.27}$$

Die zugehörigen Eigenfunktionen des Systems findet man durch Einsetzen der $\lambda_{1;2}$:

$$\begin{pmatrix} \widetilde{\vec{x}}^T(t) \\ \vec{x}^F(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{2\alpha}\left(\alpha - \Phi + 1 \pm \sqrt{R}\right) \\ 1 \end{pmatrix} \cdot e^{\lambda_{1;2} t} \tag{3.28}$$

wobei für den ersten Eigenwert die Wurzel zu addieren, für den zweiten zu subtrahieren ist.

Zur Klassifikation der möglichen Lösungen sind zwei Kriterien wichtig: Ist der Radikand R positiv, so sind beide Eigenwerte reell und man erhält exponentiell wachsende bzw. fallende Funktionen. Ist der Radikand negativ, so ergeben sich zwei linear unabhängige oszillatorische Lösungen.

Das zweite – und wichtigere – Kriterium ist die Stabilität der Lösungen: Unabhängig vom Vorzeichen des Radikanden bedeutet ein positiver Realteil des Eigenwerts eine instabile, exponentiell wachsende Lösung. Ist dagegen $Re\lambda < 0$, so ergibt sich entweder eine exponentielle Relaxation zu einer Ruhelage (negatives reelles λ) oder eine Schwingung mit exponentiell abnehmender Amplitude (komplexes λ mit negativem Realteil). Für eine vollständige Stabilität des Systems müssen selbstverständlich beide Eigenlösungen stabil sein, d.h. die Realteile beider Eigenwerte negativ sein.

Der Radikand ist negativ (und die Lösungen somit oszillatorisch), wenn genau einer der beiden Faktoren in Gl. 3.27 negativ ist, also für

$$\Phi + 1 - 2\sqrt{\Phi} \; < \; \alpha \; < \; \Phi + 1 + 2\sqrt{\Phi} \tag{3.29}$$

Diese Bedingung entspricht den innerhalb der Parabel gelegenen Bereichen II und III in Abb. 3.23. Die Lösungen aus diesem Parametergebiet oszillieren mit einer Winkelfrequenz von $\omega = \frac{1}{2}\sqrt{R}$. Über die Stabilität der oszillatorischen Lösungen entscheidet der Realteil des Eigenwertes; die Stabilitätsbedingung lautet also im oszillatorischen Fall:

$$\begin{aligned} & \alpha - \Phi - 1 && < 0 \\ \Longleftrightarrow \quad & \alpha && < \Phi + 1 \end{aligned} \tag{3.30}$$

Diese Bedingung entspricht Bereich III in Abb. 3.23. Ist Gl. 3.30 nicht erfüllt, so erhält man eine Schwingung mit exponentiell anwachsender Amplitude, d.h. eine oszillatorisch instabile Lösung (Bereich II in Abb. 3.23).

Einen positiven Radikanden und somit rein exponentielle Lösungen erhält man, wenn in Gl. 3.27 entweder beide Faktoren positiv oder beide Faktoren negativ sind. Erstere Bedingung ist erfüllt für

$$\alpha > \Phi + 1 + 2\sqrt{\Phi} \tag{3.31}$$

In diesem Fall wächst die Amplitude der Bewegung exponentiell an, ohne daß eine Korrektur möglich ist; das System verhält sich extrem instabil. In Abb. 3.23 ist dies der Bereich I.

Umgekehrt erhält man für

$$\alpha < \Phi + 1 - 2\sqrt{\Phi} \tag{3.32}$$

homogene Lösungen, die immer zur Ruhelage relaxieren. Dies ist der für den realen Betrieb als Verfolgungssystem geeignete Parameterbereich; er entspricht Bereich IV in Abb. 3.23.

Die Lösungsmenge des homogenen Gleichungssystems beschreibt das Verhaltensspektrum des Systems für $x^T(t) \equiv 0$, d.h. für ein im Nullpunkt ruhendes Target. Für den realen Betrieb ist lediglich der Parameterbereich IV (überdämpfter Fall) geeignet; die Grenzlinie zwischen den Bereichen III und IV markiert den *aperiodischen Grenzfall*.

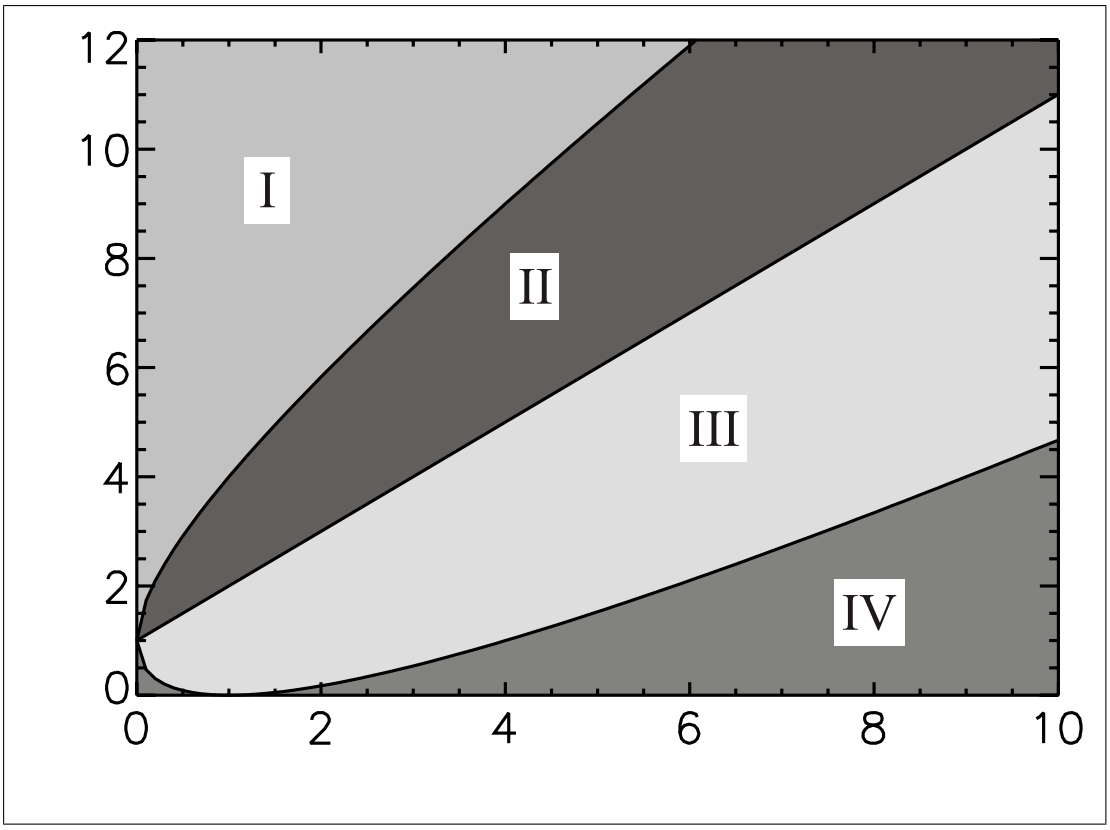


Abbildung 3.23: Verschiedene Lösungstypen der homogenen Differentialgleichung für Folgebewegungen im Parameterraum (α, Φ). Die beiden Parameter spiegeln die Eigenschaften der verwendetetn Dynamik wider: Größere Werte von Φ bedeuten eine schnellere Verfolgung von bewegtem Input durch den Aufmerksamkeitsblob; größere Werte von α bedeuten eine schnellere Nachführung der Blickrichtung bei Abweichungen. Die durch $\alpha = \Phi + 1$ beschriebene Gerade trennt stabile und instabile Lösungen, d.h. solche mit negativem bzw. positivem Realteil des zugehörigen Eigenwerts. Überlagert ist die durch $\alpha = \Phi + 1 \pm 2\sqrt{\Phi}$ gegebene Grenze zwischen oszillatorischen Lösungen mit komplexem Eigenwert und solchen mit rein reellem Eigenwert. **Bereich I:** Instabile, exponentiell wachsende Lösungen. **Bereich II:** Oszillatorische Lösungen mit exponentiell wachsender Amplitude. **Bereich III:** Oszillatorische Lösungen mit relaxierender Amplitude. **Bereich IV:** Exponentiell relaxierende Lösungen. Dieser Bereich ist der für den realen Betrieb geeignete.

3.5.8.3 Inhomogene Lösung für ein gleichmäßig bewegtes Target

Stellvertretend für beliebige Targetbewegungen $\vec{x}^T(t)$ soll in diesem Abschnitt der wichtige Spezialfall eines sich mit konstanter Geschwindigkeit bewegenden Blickziels behandelt werden. Insbesondere ist zu prüfen, ob das System eine stationäre Verfolgung leisten kann. Eine gleichmäßige Bewegung des Blickzieles mit der Geschwindigkeit $\vec{v}_0^T$ wird beschrieben durch

$$\vec{x}^T(t) = \vec{v}_0^T \cdot t \tag{3.33}$$

wobei das Koordinatensystem so gewählt ist, daß sich das Blickziel zur Zeit $t = 0$ im Ursprung befindet. Der Ansatz

$$\begin{pmatrix} \widetilde{\vec{x}}^T(t) \\ \vec{x}^F(t) \end{pmatrix} = \begin{pmatrix} \vec{v}_0^T \cdot t - \widetilde{S}^T \\ \vec{v}_0^T \cdot t - S^F \end{pmatrix} \tag{3.34}$$

entspricht einer Kameranachführung mit konstantem Schlupf S^F gegenüber dem Blickziel. $\widetilde{S}^T$ ist entsprechend der Schlupf des Aktivitätsblobs gegenüber dem Target. Einsetzen in Gl. 3.20 liefert nach wenigen Umformungsschritten die Bedingungen

$$-\Phi \cdot \widetilde{S}^T = 0 \tag{3.35}$$

$$S^F = (1-\alpha) \cdot \vec{v}_0^T \tag{3.36}$$

für die Schlupfkonstanten. Da Φ stets positiv ist, bedeutet die erste Bedingung, daß während einer stationären Verfolgung der Aktivitätsblob immer genau auf dem Target sitzt bzw. sich mit diesem mitbewegt, während die Kamera mit einem konstanten Schlupf nachgeführt wird. Diese auf den ersten Blick vielleicht überraschende Feststellung läßt sich besser verstehen, wenn man sich die Verhältnisse während einer solchen Verfolgung vergegenwärtigt: Die momentane Geschwindigkeit der Kamerabewegung ist proportional zum angezeigten Fixationsfehler $\widetilde{\vec{x}}^{Err}$, d.h. zum Abstand zwischen Kamera und dem Aktivitätsblob. Stationarität ist genau dann gegeben, wenn die so generierte Kamerabewegung genauso schnell wie die des Targets verläuft. In diesem Fall erhält der Blob (nach Gl. 3.20) exakt dieselbe Bewegungsgeschwindigkeit, wobei der Beitrag aus dem Abstand zwischen Blob und Target verschwindet. Dies ist aber konsistent mit der ersten Bedingung, wonach die Positionsdifferenz zwischen Blob und Target während der gesamten Bewegung verschwinden muß. Aus der zweiten Bedingung läßt sich ersehen, daß der dafür erforderliche Kameraschlupf S^F proportional zur Geschwindigkeit des Targets ist, d.h. schneller bewegte Ziele werden mit größerer Abweichung verfolgt. Durch Verändern von α läßt sich das genaue Verhältnis im Prinzip einstellen. In Abhängigkeit von den Anfangsbedingungen kommt zur hier betrachteten speziellen Lösung jedoch noch ein homogener Anteil hinzu – zu große Werte von α können also auch bei der Verfolgung zu einem Übersteuern der Regelschleife bis hin zur Instabilität des Systems führen.

3.5.8.4 Bewegungsprädiktion

Ein Schwachpunkt des bisher vorgestellten Systems ist die Notwendigkeit einer Abweichung zwischen Kameraposition und Blickziel zur Erzeugung einer nachführenden Kamerabewegung. Setzt sich ein ruhendes Target in Bewegung (z.B. beim Anfahren eines Autos an der Ampel), so kann das Verfolgungssystem zwangsläufig nur mit einer gewissen Verzögerung reagieren. Je nach Beschleunigung des Blickziels kann die Relaxation zu einem möglichen stationären Zustand erhebliche Zeit in Anspruch nehmen. Insbesondere nach einer Sakkade auf ein Objekt, das sich bereits bewegt, wäre aber eine sofort einsetzende Nachführung wünschenswert, um auch in diesem Fall eine zuverlässige Verfolgung zu ermöglichen.

Als Ausweg bietet sich die direkte Auswertung des vom Transientensystem gelieferten Geschwindigkeitssignals an. Dieses steht bei schnellen lokalen Bewegungen bereits nach

einem Kameraframe (also 40 ms nach Beginn der Bewegung) zur Verfügung, so daß die Nachführung sowohl früher beginnen als auch während der Bewegung besser auf Beschleunigungen reagieren kann.

Da andererseits kein Vorauseilen des Fixpunkts auftreten soll, wird das Geschwindigkeitssignal in Form einer unterschwelligen Voraktivierung der (parafovealen) Neurone der Aufmerksamkeitsschicht eingespeist. Dabei wird vorausgesetzt, daß sich das Target unter dem Fixpunkt oder zumindest in seiner Nähe befindet. (Ist dies nicht der Fall, so wird aufgrund von Gl. 3.19 eine Korrektursakkade ausgelöst.) Die *zu erwartende Targetposition* ist dann durch die Richtung der vom Transientensystem gegebenen Geschwindigkeitsantwort bestimmt; alle möglichen zu erwartenden Positionen liegen dabei in einem um den Fixpunkt zentrierten Kreisring. Die Breite diese Kreisrings (also die Ortsunsicherheit in radialer Richtung) wird dabei vom Fixationsfehler und der Unsicherheit bzw. Tuning-Unschärfe der Geschwindigkeitsdetektion bestimmt. Erregt nun jede der sechs retinotopen Geschwindigkeitsschichten einer Auflösungsstufe den in ihrer Vorzugsrichtung gelegenen Sektor des Kreisrings, so wird die zur Verfolgung des Blickziels erforderliche Wanderung des Aktivitätsblobs erleichtert und beschleunigt, und zwar gezielt in der vom Transientensystem detektierten Bewegungsrichtung. Auf diese Weise wird also eine *Bewegungsprädiktion* implementiert.

Die Geschwindigkeitsdetektoren für die verschiedenen Richtungen $\vec{v}$ sind wie folgt auf die Aufmerksamkeitsschicht aufgeschaltet:

$$w_{\vec{v}\vec{x}} = \begin{cases} h_v, & r_{min} < r_{\vec{x}} < r_{max} \quad \wedge \quad \phi_{\vec{v}} - 15^\circ < \phi_{\vec{x}} < \phi_{\vec{v}} + 15^\circ \quad \wedge \quad r_{\vec{v}} < r_0 \\ 0, & \text{sonst} \end{cases} \tag{3.37}$$

wobei $r_{\vec{x}}$ und $\phi_{\vec{x}}$ sowie $r_{\vec{v}}$ bzw. $\phi_{\vec{v}}$ Betrag und Azimut der jeweiligen Vektoren bezeichnen. Der Kreisring der zu erwartenden Targetpositionen wird also von Kreisen mit den Radien r_{min} (innen) und r_{max} (außen) begrenzt. Es wird nur der Input von Geschwindigkeitsdetektoren verwendet, die näher als r_0 am Fixpunkt liegen – eine bestehende ungefähre Fixation wird ja zu Beginn und während einer Folgebewegung vorausgesetzt. Das Gewicht h_v ist so gewählt, daß die Neurone vom Input der Geschwindigkeitsdetektoren allein nicht überschwellig werden können. Abb. 3.24 veranschaulicht die Verschaltung.

Eine alternative Möglichkeit, gezielt eine Wanderung des Aktivitätsblobs hervorzurufen, besteht im Anlegen eines gradientenbehafteten Inputs. Diese Situation wird in der Originalarbeit von Amari [1977] untersucht; es ergibt sich eine Wanderungsgeschwindigkeit, die dem Gradienten des Inputs proportional ist. Amari geht hier von einem konstanten Input-Gradient über das gesamte neuronale Feld aus. Einen solchen charakteristischen Input in Abhängigkeit von der jeweils detektierten Geschwindigkeit zu erzeugen, erscheint aufwendig, selbst wenn nur ein begrenzter Bereich um den Fixpunkt diesen Input erhalten sollte. Insbesondere wäre die notwendige Anzahl von Verbindungen im Vergleich zur hier vorgestellten Lösung um ein vielfaches höher. Aus diesen Gründen wurde diese Variante nicht weiter verfolgt.

3.5.9 Hysterese

Die bereits mehrfach erwähnte Eigenschaft, ein einmal ausgewähltes Blickziel beizubehalten, ist für eine sinnvolle Funktion der Blicksteuerung von großer Bedeutung. Würde

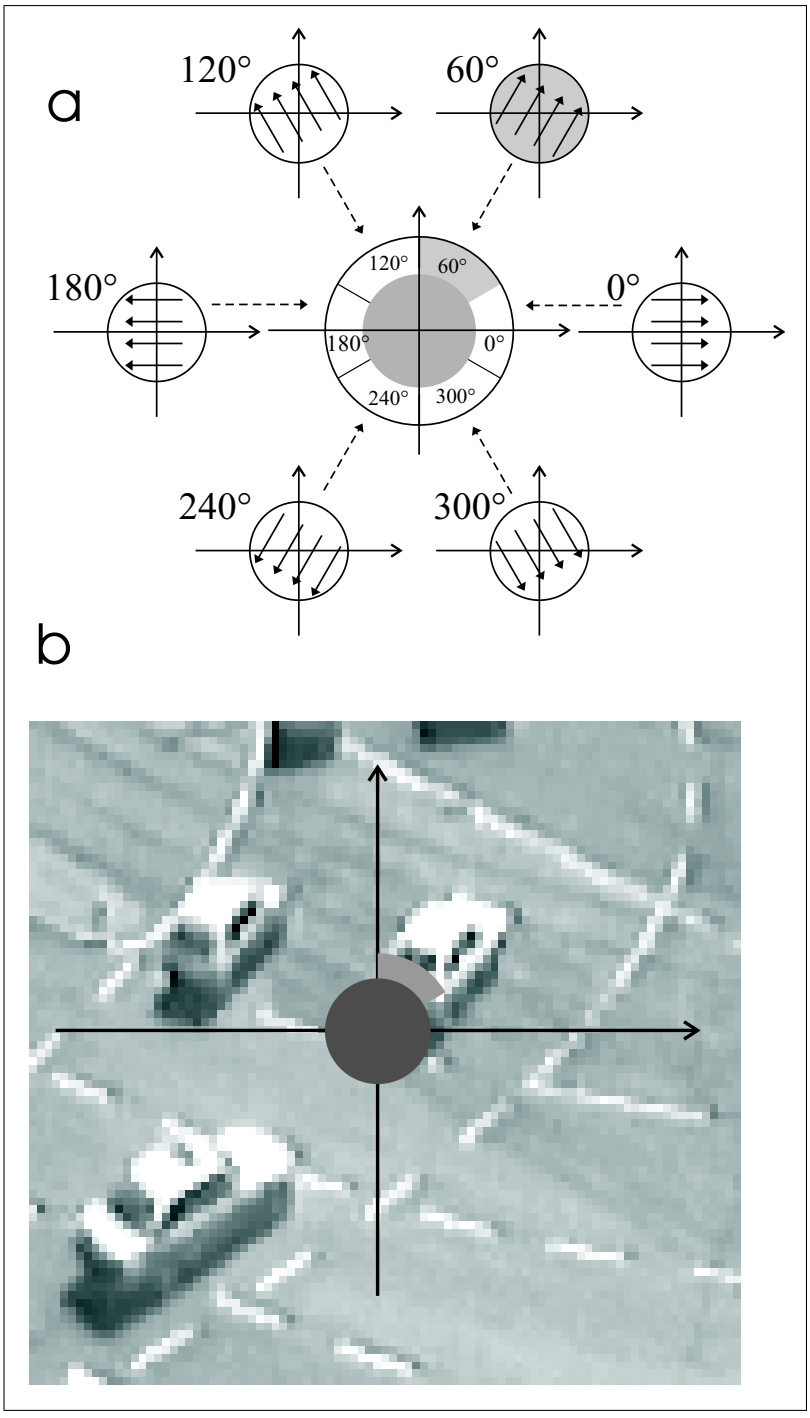


Abbildung 3.24: **(a)** Verschaltung zwischen Geschwindigkeitsdetektoren und Aufmerksamkeitsschicht zur Bewegungsprädiktion bei Folgebewegungen. Die Schaltung setzt voraus, daß das Target bereits fixiert wird, sich also im Zentrum des retinalen Koordinatensystems befindet. Dies ist z.B. nach einer Sakkade der Fall. Um die Verfolgung eines bewegten Targets zu unterstützen, wird die Aktivität der um das Zentrum gelegenen Geschwindigkeitsdetekoren ausgewertet: Jede der sechs Geschwindigkeitsschichten aktiviert entsprechend der vorhergesagten Targetposition den Sektor eines Kreisrings in der Aufmerksamkeitsschicht (gestrichelte Pfeile). Beispielhaft ist eine detektierte lokale Bewegungsrichtung von 60° hervorgehoben. Die Voraktivierung führt dazu, daß der Aktivitätsblob in die prädizierte Richtung verschoben wird und die Verfolgung bereits bei geringem Schlupf einsetzen kann. **(b)** Wirkung der Verschaltung aus Abb. a am Beispiel einer realen Szene nach einer (ungenauen) Sakkade. Der graue Kreis im Zentrum deutet die Aktivität in der Aufmerksamkeitsschicht an, der etwas hellere Sektor kennzeichnet den voraktivierten Bereich bei einer lokalen Bewegungsrichtung von 60°.

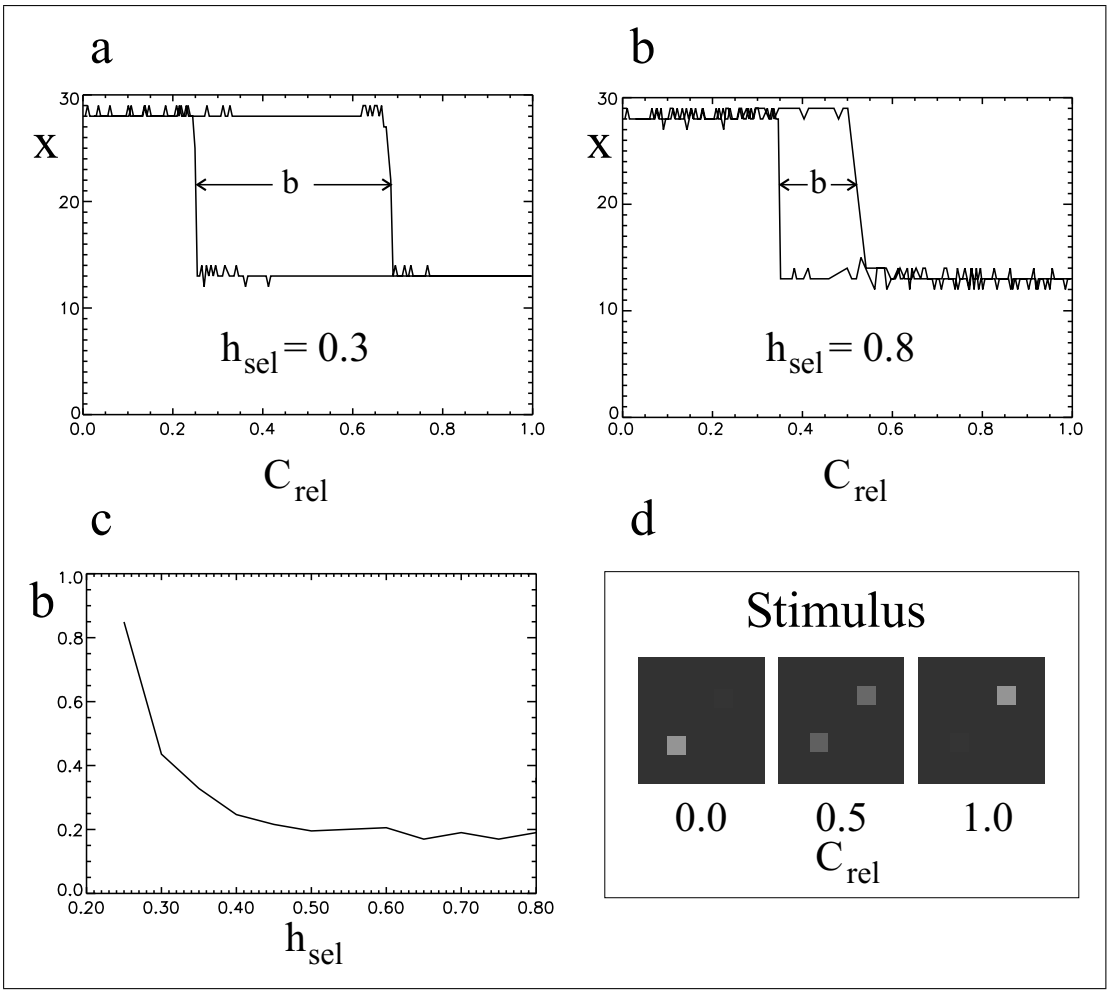


Abbildung 3.25: Hystereseschleife der Aufmerksamkeitssteuerung. Als Reiz wurde eine Folge von einfachen Grauwert-Kontrastbildern verwendet, deren relatives Kontrastverhältnis C_{rel} sich kontinuierlich vom linken zum rechten der beiden Targets (helle Quadrate) verschob. Die Stimuli für ein Kontrastverhältnis von 0, 0.5 und 1.0 sind in Abb. (d) dargestellt. (a) und (b) zeigen zwei typische Hystereschleifen für verschiedene Werte des Parameters h_{sel}. Die jeweilige Breite der Hystereseschleifen ist mit b bezeichnet und dient als ein Maß für die Stärke des Effekts. Abb. (c) zeigt die Abhängigkeit von b und h_{sel} über einen größeren Wertebereich. Bei $h_{sel} = 0.2$ tritt bei der verwendeten Reizkonstellation überhaupt kein Umspringen auf das zweite Target auf, was formal einer divergierenden Hystereseschleife entspricht.

zu jeder Zeit das jeweils stärkste Target ausgewählt, so könnten sich leicht Situationen ergeben, bei denen das System aufgrund von Rauschen oder Beleuchtungsschwankungen ständig zwischen mehreren, ähnlich auffälligen Targets hin- und herspringt. Ebenso wäre eine längerdauernde Verfolgung eines bewegten Targets kaum möglich. Andererseits widerspricht eine Hystereseeigenschaft keineswegs der geforderten schnellen Reaktion auf plötzliche Änderungen im Bild – diese müssen lediglich stark genug sein.

Mit einer aus zwei gleichartigen Targets bestehenden Reizkonfiguration, bei der der Aufmerksamkeitsinput kontinuierlich von einem Target auf das andere übergeht, läßt sich Hystereseeigenschaft quantitativ erfassen. Als Maß für das 'Beharrungsvermögen' der Aufmerksamkeitssteuerung dient dabei die Breite b der Hystereseschleife. Die Verschiebung des Inputs geschieht dabei so langsam, daß die Aufmerksamkeitsdynamik selbst quasi-

statisch arbeitet. Abb. 3.25 illustriert eine solche Simulation. Der wichtigste Parameter ist dabei der negative Offset h_{sel} der Verbindungsgewichte (s. Gl. 3.12). Je kleiner h_{sel}, desto größer ist der exzitatorische Anteil der Wechselwirkungsgewichte, d.h. derjenige Anteil der Gaußfunktion, der über der Nullinie verbleibt. Ein kleinerer Wert von h_{sel} wirkt also ähnlich wie ein stärkeres bzw. größeres Target: Die Anzahl der überschwelligen Neurone um das momentan ausgewählte Blickziel ist größer, so daß sowohl die Selbsterregung dieses Targets als auch die Inhibition der anderen Neurone insgesamt stärker ist. Dies führt im Endeffekt zu einer breiteren Hystereseschleife; diese Abhängigkeit ist in Abb. 3.25c dargestellt. Je nach Aufgabenstellung läßt sich das System durch Verändern von h_{sel} also träger oder empfindlicher einstellen.

4 Segmentierung

In diesem Kapitel wird ein theoretischer Ansatz vorgestellt, der für einen idealisierten Spezialfall eine Vorhersage von Periode und Stabilität der stationären Segmentierungsdynamik erlaubt. Die Argumentation folgt dabei weitgehend derjenigen von Gerstner et al. [1996]. Der wesentliche Unterschied liegt darin, daß in der Arbeit von Gerstner et al. lediglich vollverbundene Netze mit homogener Verschaltung betrachtet werden, wohingegen die hier verwendete Segmentierungsdynamik auf einer Netzarchitektur aufbaut, wie sie in Kap. 3.3 dargestellt wurde: eine globale Inhibition sorgt für die Phasentrennung von Objektkandidaten.

In Kap. 4.4 wird gezeigt, wie die aus der Psychophysik bekannten *aufmerksamkeitsabhängigen Latenzen* die Segmentierung beschleunigen und robuster machen können. Der Einfluß von zeitlich dispergierten Inputsignalen auf eine oszillatorische Segmentierungsdynamik wurde – unabhängig von der vorliegenden Arbeit – bereits von Opara und Wörgötter [1996] untersucht. Die Autoren beschränken sich jedoch auf elementare Simulationen mit einfachen, statischen Grauwert-Stimuli und diskutieren lediglich kontrastabhängige Latenzen. Die vorliegende Arbeit erweitert das Konzept um den Einfluß der Aufmerksamkeit und zeigt die Anwendung bei bewegten realen Szenen.

4.1 Stationäre Segmentierung

4.1.1 Stationäre Oszillationen bei exzitatorischer Kopplung

Bei überschwelliger Reizung mit konstantem Input arbeitet das Marburger Modellneuron als (quasi-)periodischer Oszillator: Abhängig von den Anfangsbedingungen wächst das Membranpotential U bis zur Feuerschwelle Θ an; beim Erreichen der Feuerschwelle wird ein Aktionspotential ausgelöst und Θ um V_Θ erhöht, was sich formal als Selbstinhibition des Neurons auffassen läßt. Dieser inhibitorische Beitrag klingt anschließend exponentiell mit der Zeitkonstante τ_Θ gegen den Schwellenoffset Θ_0 ab. Sobald das Membranpotential die abklingende Schwelle wieder übersteigt, wird der nächste Spike ausgelöst.

Die stationäre Periode dieser Oszillation läßt sich für ein einzelnes Neuron bei konstantem Input analytisch berechnen [Arndt, 1993]. Ebenso läßt sich der Ansatz auf eine Gruppe gleichartiger Neuronen übertragen, die durch exzitatorische Verbindungen mit kurzer Zeitkonstante gekoppelt sind. Arbeiten die Neuronen (z.B. aufgrund gleicher Anfangsbedingungen) völlig synchron, so hat die Kopplung keine Wirkung, da sie jeweils einen Zeitschritt nach dem Zeitpunkt des kollektiven Aktionspotentials einsetzt – ein Moment, in dem die Neuronen ohnehin refraktär sind. Eine solche völlig synchrone Oszillation muß allerdings als unrealistischer Spezialfall gelten; sowohl das Membranrauschen

der Neuronen als auch Schwankungen im realen Input lassen ein solches Verhalten im biologischen wie im technischen System nicht zu.

Trotzdem liefert die idealisierte Situation den Ausgangspunkt für die Betrachtung komplexerer Zustände: Anhand einer stationären Oszillation läßt sich ein Phasenmaß definieren, das den momentanen Zustand eines Oszillators in Bezug auf die folgende Auslösung eines Aktionspotentials wiedergibt. Zunächst soll aber das grundlegende, in der vorliegenden Arbeit verwendete Segmentierungsnetz mit seiner Dynamik vorgestellt werden.

4.2 Segmentierung mit globaler Inhibition

In Kap. 3.3 wurde bereits das Konzept vorgestellt, vermutlich zusammengehörige Bildelemente durch *Linking*-Verbindungen zu koppeln, und so die jeweiligen sie repräsentierenden Neurone in ihrer Aktivität zu synchronisieren. Da neben der Zusammengehörigkeit von Bildorten aber auch die Trennung von Objekten codiert werden soll, erweiterten Arndt et al. [1992] die exzitatorische Kopplungen um eine inhibitorische Komponente, die gleichmäßig zwischen *allen* Neuronen wirkt. Diese inhibitorische Komponente wird von der gesamten Netzwerkaktivität gespeist und wirkt auch wieder auf das gesamte Netz zurück, weshalb sie auch als *globale Inhibition* bezeichnet wird. Im Verein mit der *lokal* wirkenden exzitatorischen Kopplung können sich so Neuronenverbände herausbilden, deren Aktivität untereinander synchronisiert, gegeneinander jedoch aufgrund der gegenseitigen Inhibition phasenversetzt ist. Die globale Inhibition sorgt dabei für einen temporären Winner-Take-All-Wettbewerb, bei dem jeweils nur einer der Neuronenverbände auf einmal aktiv ist. Abb. 4.1 zeigt die zugehörige neuronale Verschaltung, Abb. 4.2 das typische zugehörige Aktivitätsmuster.

Den vollständigen Zustandsraum einer solchen gegenphasigen Oszillation analytisch zu beschreiben, erweist sich als schwierig, weil der Schwellenvergleich eine starke Nichtlinearität darstellt. Eine Formulierung in einem System gekoppelter Differentialgleichungen ist zwar prinzipiell möglich, führt aber aufgrund der großen Neuronenzahl und der darin vorkommenden nicht-stetigen Funktionen nicht viel weiter (zu einer Näherung s. [Götzl, 1994]).

Andererseits ist das Verhalten des Systems zwischen den Aktionspotentialen vollständig darstellbar, da hier lediglich das Abklingen von Tiefpässen zu berücksichtigen ist. Voraussetzung für die rechnerische Ausnutzung dieser Eigenschaft ist also ein hinreichend großer zeitlicher Abstand zwischen aufeinanderfolgenden Aktionspotentialen. Dies läßt sich weitgehend sicherstellen, indem alle inhibitorischen Zeitkonstanten deutlich größer als die exzitatorischen gewählt werden. Zu den inhibitorischen Zeitkonstanten zählt neben der Zeitkonstante der globalen Inhibition, τ_I, auch diejenige der Refraktärwirkung, τ_Θ. Die Zeitkonstanten der exzitatorischen Tiefpässe sind τ_F für das Feeding und τ_L für das Linking. Nimmt man nun an, daß das Netz sich bereits in einem eingeschwungenen Zustand mit phasenversetzter Aktivität mehrerer Neuronengruppen befindet, so läßt sich über einen *Selbstkonsistenzansatz* die mögliche Periode der gesamten Oszillation abschätzen. Außerdem läßt sich zeigen, daß ein solcher Zustand lokal stabil ist, d.h. kleine Abweichungen davon werden von der Dynamik selbst ausgeglichen.

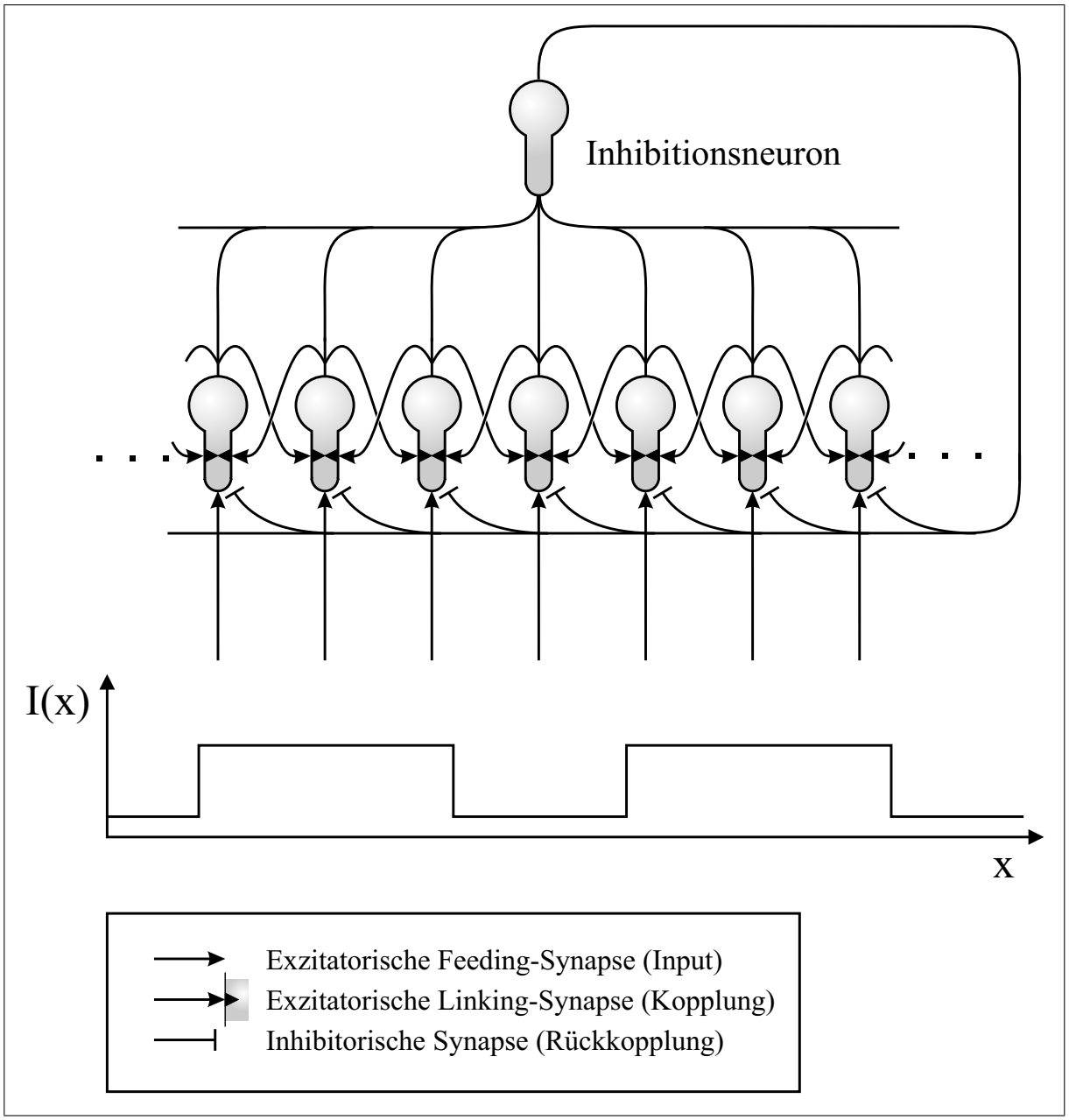


Abbildung 4.1: **(a)** Eindimensionales Segmentierungsnetzwerk zur Objektseparation im Zeitbereich. Benachbarte Neurone sind über die Linking-Verbindungen exzitatorisch gekoppelt und können so bei ähnlicher Anregung ihre Spikeaktivität synchronisieren. Die Reichweite dieser Verbindungen ist im Regelfall größer als 1 Neuron (gezeichnet). Die als *globale Inhibition* bezeichnete Rückkopplungsschleife führt zusätzlich zu einer temporären gegenseitigen Hemmung aller im Netz vorhandenen Neuronengruppen: Ist eine größere Gruppe von Neuronen gleichzeitig aktiv, so wird dadurch das gesamte Netzwerk solange gehemmt, bis die Inhibition wieder abgeklungen ist. **(b)** Beispielhaftes eindimensionales zu segmentierendes Inputprofil $I(x)$. Ist die Lücke zwischen den beiden Teilbereichen größer als die Reichweite der Linking-Verbindungen, so kann aufgrund des von der globalen Inhibition induzierten temporären Winner-Take-All-Wettbewerbs eine zeitliche Trennung der Spikeaktivität von linkem und rechtem Netzwerkbereich erfolgen. Dabei wird die Aktivität der Einzelbereiche zusätzlich synchronisiert, so daß die Zusammengehörigkeit von Teilen des Intensitätsprofils im zeitlichen Verlauf der Netzwerkaktivität codiert wird. Die zugehörige Netzwerkaktivität nach erfolgter Segmentierung zeigt Abb. 4.2.

4.2.1 Bestimmung der Oszillationsperiode im stationären Zustand

Um die Oszillationsperiode im stationären Zustand näherungsweise analytisch zu bestimmen, sind einige Vereinfachungen und Näherungen notwendig. Diese sind im einzelnen:

1. Die beteiligten Neurone erhalten denselben statischen, überschwelligen Input F_0.

2. Das Netzwerk befindet sich bereits im eingeschwungenen Zustand, d.h. seine Aktivität ist seit unendlich langer Zeit periodisch.

3. Eine Anzahl von Objekten n ist durch relativ zueinander phasenversetzte Aktivität der sie repräsentierenden Neuronengruppen codiert.

4. Diese Neuronenverbände arbeiten in sich völlig synchron.

5. Die relative Phasenverschiebung der Objekte gegeneinander beträgt einen Bruchteil $1/n$ der gesamten Oszillationsperiode.

6. Im Verlauf einer Periode ist jede Neuronengruppe genau einmal aktiv; die Aktivierung geschieht in zyklischer Folge. Bei jeder Aktivierung eies Neuronenverbandes wird das globale Inhibitionsneuron genau einmal aktiv, insgesamt also n mal in jeder Periode.

7. Die Dauer für eine Rückwirkung der inhibitorischen Einflüsse (Schwelle und globale Inhibition) ist vernachlässigbar klein gegenüber der Oszillationsperiode und wird zu Null angenommen.

8. Für die Zeitkonstanten der Tiefpässe gilt folgende Relation:

$$\tau_L \ll \tau_F < \tau_I < \tau_\Theta \tag{4.1}$$

9. Für die Verstärkungsfaktoren der inhibitorischen Tiefpässe gilt:

$$V_\Theta > V_I \tag{4.2}$$

Während die Relationen zwischen den Zeitkonstanten leicht einzustellen sind, stellen die ersten sieben Annahmen starke Idealisierungen dar; die Punkte zwei und sieben können beide nicht streng erfüllt sein: Weder existiert ein reales Netz seit unendlicher langer Zeit, noch gibt es in realen Netzen eine beliebig schnelle Rückkopplung. Dennoch erlauben diese Annahmen eine Analyse der Funktionsweise der Segmentierungsdynamik und geben Hinweise, wie diese in realen Netzen robuster zu gestalten ist.

Die letzte Forderung stellt sicher, daß die Kopplung der Neurone untereinander zwar einerseits einen Einfluß auf den Zeitpunkt des nächsten Aktionspotentials eines Neurons haben kann, andererseits das Potential an der Linking-Synapse so schnell abklingt, daß sein Beitrag zum Membranpotential bei der erneuten Auslösung eines Spikes vernachlässigt

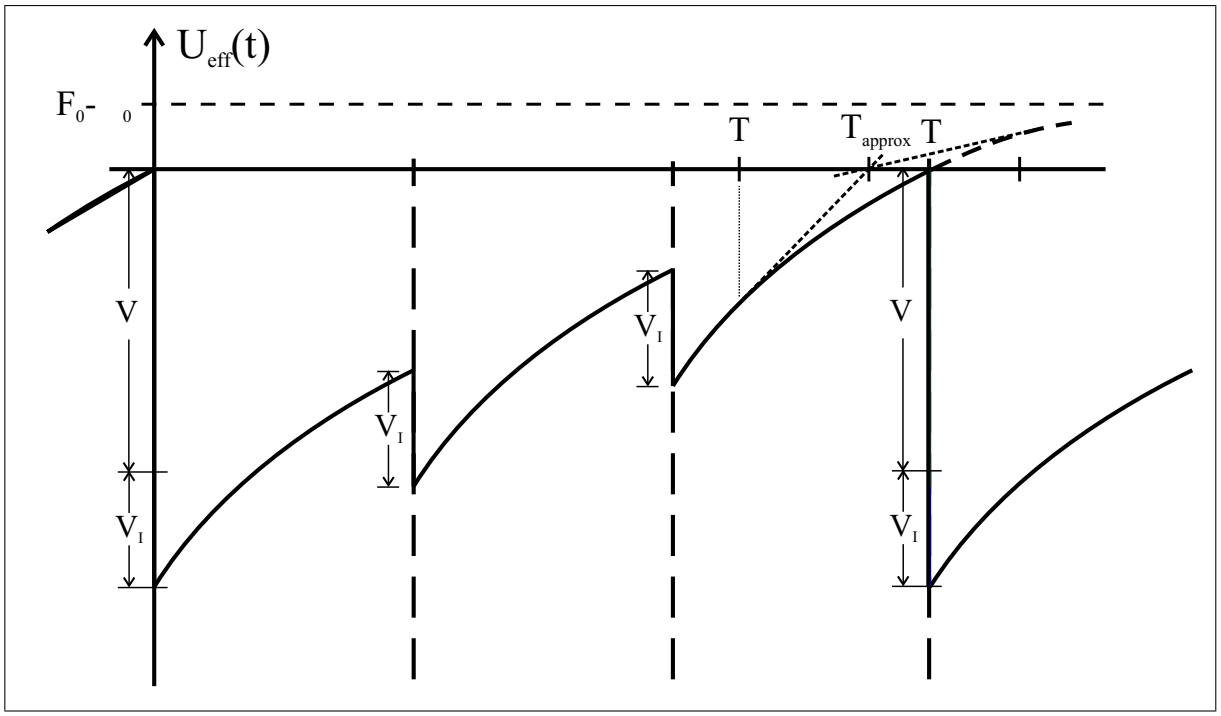


Abbildung 4.2: Ideales Aktivitätsmuster in einem Segmentierungsnetz mit globaler Inhibition und drei separierten Objekten. Alle Neurone, die ein Objekt repräsentieren, feuern gleichzeitig. Die Aktivität der jeweils anderen Neuronenverbände (die zu einem davon getrennten Objekt gehören), ist dagegen phasenverschoben. Voraussetzung für das Zustandekommen einer solchen Dynamik sind ausreichend lange inhibitorische Zeitkonstanten und geeignete Anfangsbedingungen.

werden kann. Lange inhibitorische Zeitkonstanten sind erforderlich, um bei Schwankungen im Input eine Überlappung der Spikeauslösung (und damit einen Zusammenbruch der Objektseparation) zu vermeiden.

In Abb. 4.2 ist eine Periode einer solchen Oszillation am Beispiel von $n = 3$ Objekten dargestellt. Aufgetragen ist der Zeitverlauf des *effektiven Membranpotentials* U_{eff} aller Neurone einer Gruppe (s. Kap. 2.6). Erreicht dieses Null, so überwiegt die Summe der exzitatorischen Beiträge zum Membranpotential diejenige der inhibitorischen, und ein Aktionspotential wird ausgelöst. Die Zeitachse ist so gewählt, daß ihr Nullpunkt mit einem Aktionspotential des betrachteten Neurons zusammenfällt. Das Ende der folgenden Periode – und damit ihre Dauer – läßt sich nun aus dem Zeitpunkt bestimmen, zu dem der nächste Spike ausgelöst wird. Dabei kann man ausnutzen, daß zwischen den Aktionspotentialen alle Tiefpässe ungestört abklingen. Der inhibitorische Tiefpaß wird bei jeder Aktivität einer anderen Neuronengruppe um V_I aufgeladen, im Verlauf einer Periode also $n-1$ mal. Die n-te Rückwirkung der globalen Inhibition fällt wieder mit dem nächsten Aktionspotential des betrachteten Neurons zusammen. Die Linking-Kopplung spielt in der Rechnung keine Rolle, da vorausgesetzt wurde, daß die Neuronengruppen untereinander bereits vollständig synchronisiert sind.

Für das effektive Membranpotential ergibt sich also folgender Zeitverlauf:

$$\begin{aligned} U_{\mathit{eff}}(t) &= F(t) \cdot (1 + L(t)) - \Theta(t) - I(t) \\ &\approx F_0 - \Theta_0 - \Theta_{dyn}(t) - I(t) \end{aligned} \tag{4.3}$$

wobei die Bezeichnungen denen aus Kap. 2.6 entsprechen.

Der *Selbstkonsistenzansatz* beruht nun darauf, daß der Zeitverlauf der Teilpotentiale in jeder Periode einerseits exakt mit dem in der vorhergegegangenen Periode identisch ist, andererseits das Netzwerk bereits seit unendlich langer Zeit aktiv ist und somit jedes Teilpotential als Überlagerung aller identischen vorangegangen Perioden darzustellen sein soll. In der Praxis spielen die Terme, die weit zurückliegenden Perioden entsprechen, natürlich keine Rolle mehr. Für die Rechnung ist die Formulierung der Teilpotentiale als unendliche Reihen jedoch nützlich.

Der dynamische Anteil der Schwelle errechnet sich folgendermaßen:

$$\begin{aligned} \Theta_{dyn}(t) &= \sum_{k=0}^{\infty} V_\Theta \cdot e^{\frac{t+kT}{\tau_\Theta}} \\ &= V_\Theta^\infty \cdot e^{-\frac{t}{\tau_\Theta}} \qquad \text{mit} \qquad V_\Theta^\infty = \frac{V_\Theta}{1 - e^{\frac{T}{\tau_\Theta}}} \end{aligned} \tag{4.4}$$

Jeder Summand repräsentiert dabei den Beitrag aus einer Periode (um $-k \cdot T$ verschobene Stoßantwort des Tiefpaß). V_Θ^∞ wurde nach Ausklammern mit Hilfe der Formel für unendliche geometrische Reihen bestimmt.

Für die globale Inhibition verläuft die Rechnung analog. Allerdings ist hier zu berücksichtigen, daß das Inhibitionsneuron während jeder Periode n-mal aktiv ist, wobei jedesmal alle Neurone im Netz einen Inhibitionsimpuls erhalten. Der Laufindex k zählt also in der folgenden Gleichung keine Perioden, sondern Aktionspotentiale des Inhibitionsneurons:

$$\begin{aligned} I(t) &= \sum_{k=0}^{\infty} V_I \cdot e^{-\frac{t+kT/n}{\tau_I}} \\ &= V_I^\infty \cdot e^{-\frac{t}{\tau_I}} \qquad \text{mit} \qquad V_I^\infty = \frac{V_I}{1 - e^{\frac{T}{\tau_I}}} \end{aligned} \tag{4.5}$$

wobei V_I^∞ wieder mit Hilfe der Formel für die unendliche geometrische Reihe bestimmt wurde.

Aus dieser Darstellung ergibt sich, daß sich das Verhalten eines seit unendlich langer Zeit aktiven Neurons formal auf ein Neuron mit der Schwellenverstärkung V_Θ^∞ und der Inhibitionsverstärkung V_I^∞ abbilden läßt, das zum Zeitpunkt $t = 0$ zum ersten Mal gefeuert hat. Die Auslösung des nächsten Aktionspotentials soll nun genau am Ende der Periode, also bei $t = T$ erfolgen. Die Bedingung dafür lautet also:

$$0 = U_{eff}(T) = F_0 - \Theta_0 - V_I^\infty \cdot e^{-\frac{T}{\tau_I}} - V_\Theta^\infty \cdot e^{-\frac{T}{\tau_\Theta}} \tag{4.6}$$

Diese Gleichung ist nur im Spezialfall $\tau_\Theta = \tau_I$ geschlossen lösbar. Eine Näherung läßt sich aber durch eine Reihenentwicklung gewinnen. Um einen sinnvollen Entwicklungspunkt zu finden, kann man zunächst $\tau_I = \tau_\Theta$ setzen, d.h. man berechnet diejenige Periode T_Θ, die sich allein aufgrund der Schwellenzeitkonstante ergäbe. Diese ist

$$T_\Theta = -\tau_\Theta \cdot ln\left(\frac{F_0 - \Theta_0}{V_\Theta^\infty + V_I^\infty}\right) \tag{4.7}$$

Zur übersichtlicheren Notation substituieren wir das Argument des Logarithmus:

$$\beta := \frac{F_0 - \Theta_0}{V_\Theta^\infty + V_I^\infty} \qquad \text{sowie} \qquad \gamma := \frac{\tau_\Theta}{\tau_i} \tag{4.8}$$

Es gelten dann die Relationen

$$e^{-\frac{T_\Theta}{\tau_\Theta}} = \beta^{-1} \qquad \text{sowie} \qquad e^{-\frac{T_\Theta}{\tau_I}} = \beta^{-\gamma} \tag{4.9}$$

Damit ergibt die Reihentwicklung von $U_{eff}(t)$ um T_Θ

$$\begin{aligned} U_{eff}(t) &= U_{eff}(T_\Theta) + \left.\frac{dU_{eff}}{dt}\right|_{t=T_\Theta} \cdot (t - T_\Theta) + \ldots \\ &\approx F_0 - \Theta_0 - V_\Theta^\infty \beta^{-1} - V_I^\infty \beta^{-\gamma} + \left(\frac{V_\Theta^\infty}{\tau_\Theta}\beta^{-1} + \frac{V_I^\infty}{\tau_I}\beta^{-\gamma}\right) \cdot (t - T_\Theta) \end{aligned} \tag{4.10}$$

Einsetzen in Gl. 4.6 und auflösen nach T liefert:

$$\begin{aligned} T &= T_\Theta + \frac{V_\Theta^\infty \beta^{-1} + V_I^\infty \beta^{-\gamma} + \Theta_0 - F_0}{\frac{V_\theta^\infty}{\tau_\Theta}\beta^{-1} + \frac{V_I^\infty}{\tau_i}\beta^{-\gamma}} \\ &= -\tau_\Theta \cdot \left(-ln\beta + 1 + \frac{(1-\gamma)\beta^{1-\gamma} - \frac{(F_0-\Theta_0)^2}{(V_\Theta^\infty + V_I^\infty)V_I^\infty}}{\frac{V_\Theta^\infty}{V_I^\infty} + \gamma\beta^{1-\gamma}} \right) \end{aligned} \tag{4.11}$$

Aufgrund der rechtsgekrümmten Aufladekurve des Membranpotentials ist die so erhaltene Näherung für T immer zu klein; die in Abb. 4.2 eingezeichneten Tangenten illustrieren den Effekt der Reihentwicklung um T_Θ.

4.2.2 Stabilität und Minimumseigenschaft der stationären Oszillationen

Die bis jetzt angestellten Betrachtungen liefern zwar eine Abschätzung für die Oszillationsperiode im stationären Zustand, beantworten aber noch nicht die Frage nach der Stabilität dieses Zustandes bzw. ob und von welchen Anfangsbedingungen aus er überhaupt erreicht werden kann.

Um einen genauen Eindruck von der Arbeitsweise der Synchronisations- und Separationsdynamik zu bekommen, ist es sinnvoll, den Mechanismus, der zur dargestellten Objektseparation führt, im Detail zu untersuchen. Dabei spielt – wie bei der Abschätzung der Oszillationsperiode – die Frage nach dem nächsten Spikezeitpunkt eines Neurons die entscheidende Rolle. Die Zeit, die das Neuron ohne das Eintreffen weiterer Spikes von anderen Neuronen bis zur Auslösung seines nächsten Aktionspotentials benötigen würde, eignet sich als eine Art Phasenmaß, das den Fortgang der neuronalen Anregung beschreibt. Eine strenge Definition der Phase analog derjenigen einer harmonischen Schwingung ist ohne weiteres nicht möglich, weil diese eine bekannte, feste Aktivitätsperiode voraussetzt.

Zur Entstehung der Separationsdynamik tragen im wesentlichen drei Teilprozesse bei:

1. Die exzitatorische Linking-Kopplung zwischen Neuronen mit potentiell zusammengehörigen rezeptiven Feldern führt zu einer Annäherung der Spikezeitpunkte der Neurone. Das in der Phase vorlaufende Neuron feuert zuerst und bringt durch den zusätzlichen Exzitationspuls die benachbarten Neurone dichter an die Feuerschwelle. Die Auswirkung dieser Exzitation ist in Abb. 4.3 dargestellt.

2. Neurone, deren Membranpotential noch weiter von der Schwelle entfernt sind, erhalten von der globalen Inhibition einen inhibitorischen Impuls, der mit der relativ langen Zeitkonstante τ_I nachwirkt. Solche Neurone werden also in ihrer Phase zurückgesetzt (Abb. 4.4a). Dieser Effekt ist um so stärker, je näher das Neuron bereits an der Schwelle war. Dadurch werden die Phasen aller momentan nicht aktiven Neurone im Netz einander in hohem Maß angeglichen (Abb. 4.4b).

3. Neurone, die ohne die globale Inhibition kurz nach dem vorlaufenden Neuron gefeuert hätten, sind nur dann von der Phasenrücksetzung ausgenommen, wenn sie noch vor Eintreffen des Inhibitionspulses feuern. Dies ist immer dann der Fall, wenn ihre Phase der des vorlaufenden Neurons bereits so ähnlich ist, daß der exzitatorische Kopplungspuls genügt, um sie innerhalb von 1–2 Zeitschritten ebenfalls zum Feuern zu bringen.

In der Summe führen diese Teilprozesse dazu, daß Neurone, die ähnlichen Input erhalten und durch ausreichend starke Kopplung verbunden sind, ihre Spikezeitpunkte synchronisieren, während sie von denen aller anderen Neuronen im Netz zeitlich getrennt werden. Die rhythmische Aktivität der globalen Inhibition führt dabei dazu, daß regelrechte Zeitschlitze entstehen, in denen Neuronengruppen aktiv sein können. Eine Voraussetzung dafür ist die rechtsgekrümmte Aufladekurve des Membranpotentials; diese ist auch für die Stabilität des gesamten Systems von entscheidender Bedeutung [Ernst, 1993].

Eine vergleichbare Argumentation für den Fall einer ebenfalls impulscodierten, aber in ihrer räumlichen Wirkung begrenzten Inhibition geben auch Nischwitz und Glünder [1995]. Aufgrund der lokalen Natur der rückwirkenden Inhibition ergeben sich in ihrem Modell synchronisierte Teilverbände von Neuronen, ohne daß zwischen diesen eine globale Phasenbeziehung besteht – was der physiologischen Realität sicherlich näher kommt. Zur Erklärung der Synchronisation innerhalb der Neuronenverbände ziehen Nischwitz und Glünder jedoch i.w. die gleichen Mechanismen heran, die oben vorgestellt wurden.

4.3 Definition eines Segmentierungsmaßes

Ausgehend vom symmetrischen stationären Zustand wie er in Abb. 4.2 dargestellt ist, läßt sich anhand der Kreuzkorrelation zweier Aktivitätsmuster ein einfaches Maß für den Fortschritt der Segmentierung in einer realen Dynamik definieren. Der Wert 1 soll dabei für eine vollständige Segmentierung stehen, der Wert 0 synchrone Aktivität *verschiedener* Objekte, also fehlende Segmentierung, anzeigen. Abb. 4.5 illustriert die dabei involvierten Größen.

Zunächst muß dazu der Zeitverlauf der neuronalen Aktivierung wie folgt analysiert werden:

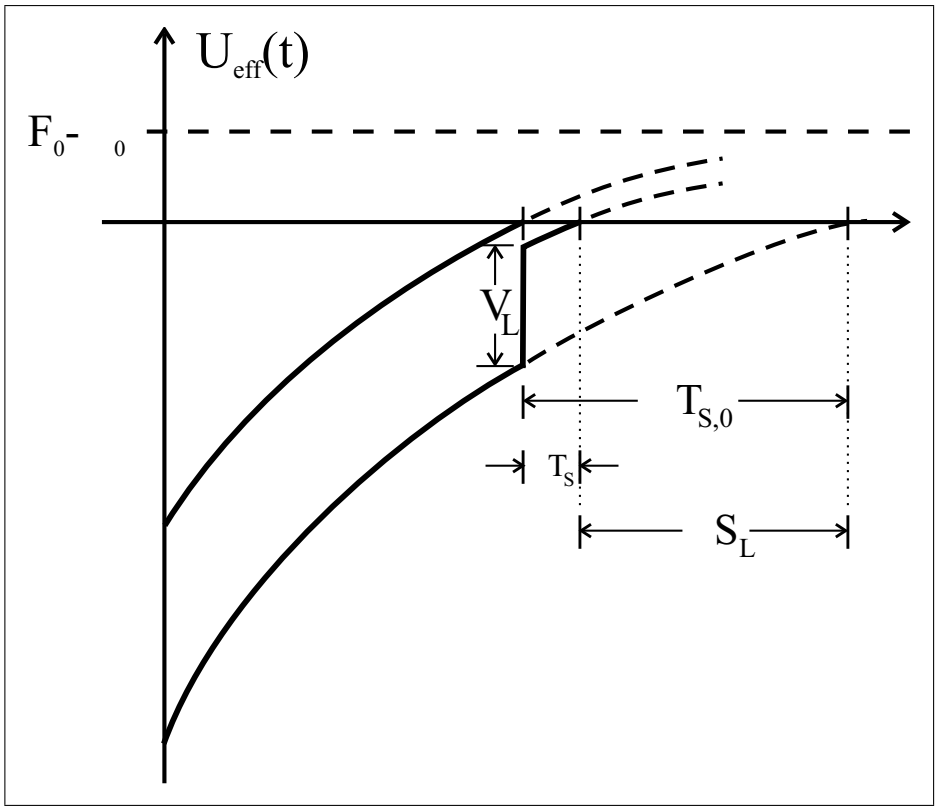


Abbildung 4.3: Auswirkung der Linking-Kopplung auf den Spikezeitpunkt der Neurone. Das vorlaufende Neuron feuert zuerst und schiebt damit das nachfolgende Nachbarneuron durch einen Linking-Puls in der Phase nach vorne. Die urpsrüngliche Phasendifferenz $\Delta T_{S,0}$ der beiden Neurone wird dadurch um ΔS_L auf ΔT_S reduziert.

1. Zuordnung aller Spikes zu Objekten (bzw. dem Szenenhintergrund) entsprechend den RFs der Neurone

2. Zusammenfassen der Spikes zu den Objekten bzw. dem Hintergrund gehörenden Massensignalen (*Multi-Unit-Activity*, MUA)

$\Rightarrow$ Zu jedem Objekt sowie ggf. dem Hintergrund entsteht ein Massensignal.

3. Paarweise Berechnung der Kreuzkorrelation der Massensignale

Anhand der Kreuzkorrelation läßt sich nun gleichzeitige bzw. phasenversetzte Aktivität der verschiedenen Neuronengruppen feststellen: Ein Peak bei Null in der Kreuzkorrelation zeigt gleichzeitige Aktivität an, ein Peak bei der Hälfte der Oszillationsperiode T dagegen phasenversetzte Aktivität. Voraussetzung für die weitere Rechnung ist, daß tatsächlich getrennte Peaks vorliegen. Diese Bedingung ist genau dann erfüllt, wenn sich eine Dynamik vom oben beschriebenen Typ eingestellt hat. Ist diese Voraussetzung nicht erfüllt, dann ist auch die Anwendung des vorgestellten Segmentierunsmaßes unsinnig. Bezeichnen wir die zu zwei Objekten gehörenden Massensignale mit $MUA_1(t)$ und $MUA_2(t)$, so ist die Kreuzkorrelation wie folgt definiert:

$$CC_{21}(t) = MUA_1(t) \star MUA_2(t) = \int_{-\infty}^{\infty} MUA_1(\tau) MUA_2(t+\tau) d\tau \qquad (4.12)$$

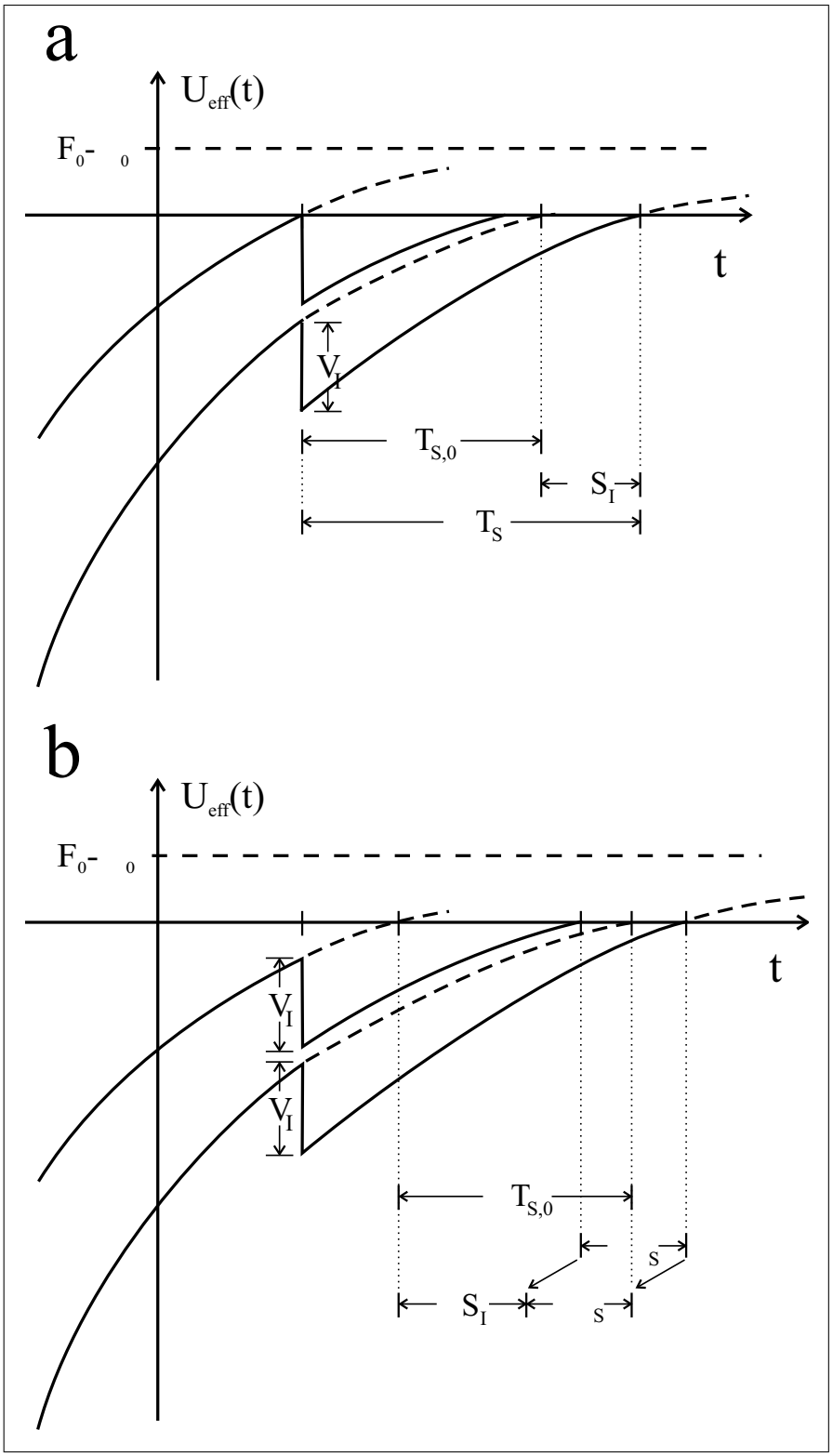


Abbildung 4.4: Wirkung der globalen Inhibition auf die relative Phase von zwei Neuronen. **(a)** Desynchronisierende Wirkung. Das vorlaufende Neuron löst die globale Inhibition aus und inhibiert so indirekt das nachlaufende. Die relative Phasendifferenz $\Delta T_{S,0}$ der beiden Neurone wird dadurch stark vergrößert, die Neurone somit desynchronisiert. **(b)** Synchronisierende Wirkung. Eine dritte Neuronengruppe (nicht gezeichnet) löst die globale Inhibition aus und inhibiert beide betrachtete Neurone. Das vorlaufende Neuron wird im Vergleich zum nachlaufenden stärker in der Phase zurückgeworfen, so daß sich die relative Phasendifferenz verringert: Die beiden Neurone werden synchronisiert.

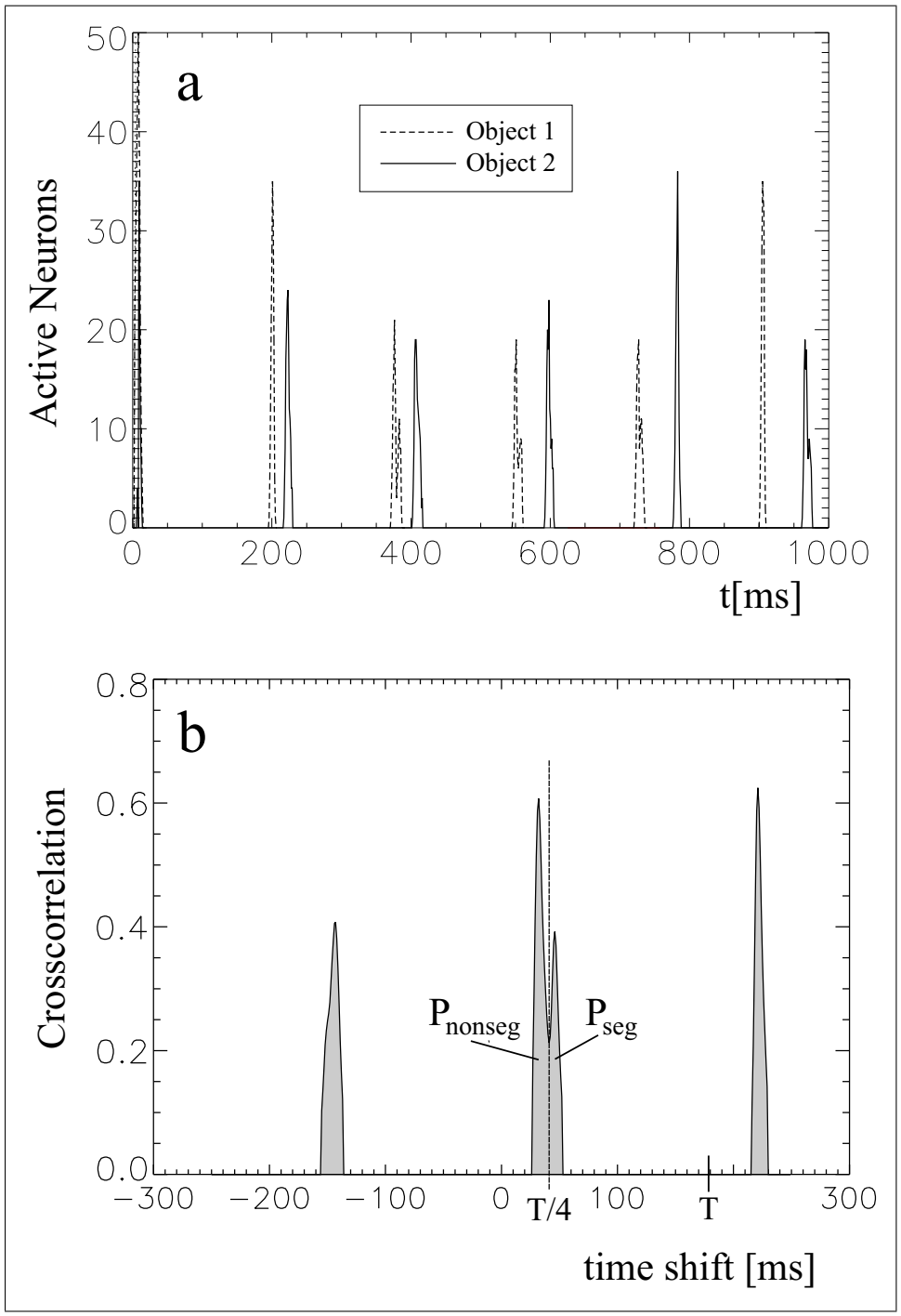


Abbildung 4.5: Zur Definition des Segmentierungsindex SI. **(a)** Oszillatorische Massensignale zweier Neuronengruppen, die jeweils ein zu segmentierendes Objekt repräsentieren. Die Dynamik entspricht der aus Abb. 4.2; hier sind lediglich die Ausgangssignale aufgetragen. **(b)** Aus den beiden Zeitverläufen wird nach Gl. 4.12 eine zeitaufgelöste Kreuzkorrelation berechnet. Der gezeigte Verlauf der Kreuzkorrelation entspricht dem hervorgehobenen Datenpunkt in Abb. 4.6. In der vorliegenden, einfachen Grundlagensimulation tritt lediglich am Anfang tatsächlich überlappende Aktivität der Neuronengruppen auf; während der weiteren Simulation wandert der Peak von Null weg und zeigt so die wachsende Phasendifferenz der beiden Neuronengruppen an. Als Grenze wird nach Gl. 4.14 bzw. Gl. 4.13 eine Phasenverschiebung von einem Viertel der stationären Periode T verwendet.

Die Fläche P_{nonseg} unter dem Peak bei Null ist:

$$P_{nonseg} = \int_{-\frac{T}{4}}^{\frac{T}{4}} CC_{21}(t)dt \tag{4.13}$$

Entsprechend ergibt sich der segmentierte Anteil P_{seg} der Aktivität als Fläche unter dem Peak bei $\frac{T}{2}$:

$$P_{seg} = \int_{\frac{T}{4}}^{\frac{3}{4}T} CC_{21}(t)dt \tag{4.14}$$

Der Segmentierungindex SI wird nun definiert als:

$$SI = 1 - \frac{P_{nonseg}}{P_{seg}} \tag{4.15}$$

4.4 Segmentierung mit Latenzen

4.4.1 Segmentierung einfacher Reize

In Kap. 2.5.2 wurde bereits dargestellt, daß der Input des natürlichen Sehsystems bereits stark über die Zeit verteilt ist. Im Sinn einer zeitlichen Szenensegmentierung kann man also davon sprechen, daß das Kontur-Form-System bereits *vorsegmentierten* Input erhält. Das Segmentierungsnetz muß in diesem Fall 'nur noch' die im Input-Spikestrom vorhandenen zeitlichen Unterschiede aufrechterhalten bzw. entsprechend der räumlichen Zugehörigkeit der Objekte umordnen. Stimmt die Vorsegmentierung im Input bereits gut mit der endgültigen Anordnung überein, so gelangt das Netzwerk dadurch wesentlich schneller in die Nähe eines stationären Zustandes als das bei gleichzeitig eintreffendem Input der Fall wäre. Probleme können auftreten, wenn sich kontrastinduzierte Latenzen mit den aufmerksamkeitsbedingten überlagern, etwa bei ungleichmäßiger Beleuchtung.

In der o.a. Arbeit von Opara und Wörgötter wird dieser Effekt auf ganz ähnliche Weise wie in der vorliegenden Arbeit untersucht, wobei sich ihr Segmentierungsnetz vom hier verwendeten etwas unterscheidet. Trotzdem stimmen ihre Ergebnisse für einfache Reize (Rechtecke unterschiedlichen Grauwerts) sehr gut mit den im folgenden Abschnitt vorgestellten überein.

Abb. 4.6 zeigt die Abhängigkeit des Segmentierungsfortschritts von der zeitlichen Dispersion der beiden Objekte am Eingang des Netzwerks. Bis zu einer Zeitdifferenz, die der stationären Oszillationsperiode entspricht, beschleunigt eine eingangsseitige Dispersion die Entstehung der endgültigen Segmentierungsdynamik.

4.4.2 Implikationen für die Segmentierung realer Szenen

Wie sich in Kap. 5.2 zeigen wird, wird die Segmentierungsdynamik durch die Dispersion jedoch nicht nur schneller, sondern auch robuster. Das geht so weit, daß bei realen Anwendungen mit ihren komplexen Intensitätsmustern eine quasi-stationäre Segmentierung oftmals nur gelingt, wenn im Inputstrom des Netzwerks bereits eine zeitliche Dispersion vorliegt. Zudem kann eine solche Vorgsegmentierung auch dazu verwendet werden, einem

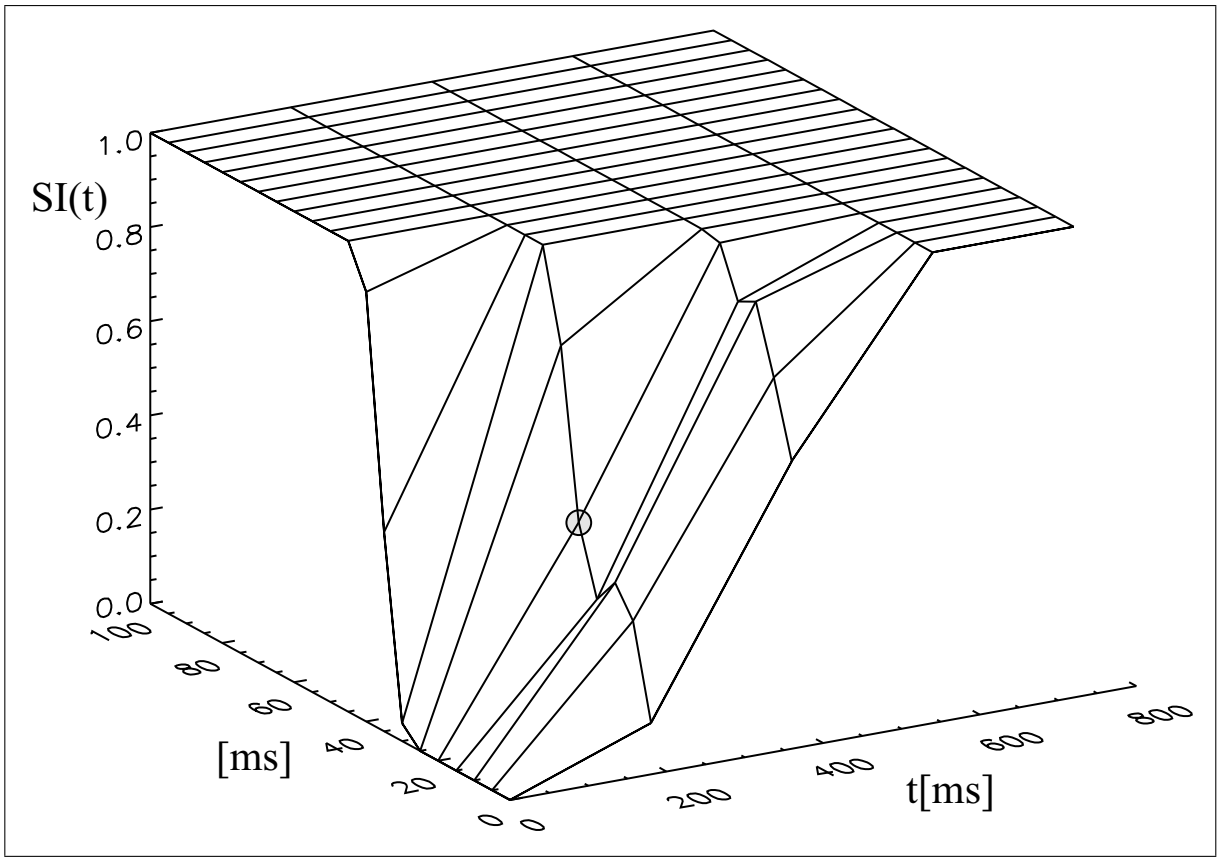


Abbildung 4.6: Separation zweier Rechtecke mit unterschiedlicher Eingangslatenz. Die beiden Objekte werden mit einem festen anfänglichen Zeitversatz Δ auf das Segmentierungsnetz gegeben. Auf der Ordinate ist der Segmentierungsindex SI aufgetragen. Der stationäre, segmentierte Zustand ($SI = 1$) wird in allen Fällen erreicht. Ausgehend von $\Delta = 0$ (kein Zeitversatz) gelingt die Segmentierung jedoch um so früher, je größer der anfängliche Latenzunterschied ist. Die in Abb. 4.5b beispielhaft gezeigte Kreuzkorrelation wurde am eingekreisten Datenpunkt ermittelt.

nachgeschalteten Assoziativspeicher zeitlich getrennte Muster zu präsentieren, ohne daß eine Oszillation im eigentlichen Sinn notwendig wäre. Prinzipiell läßt sich mit diesem Verfahren also eine *One-Shot-Analyse* einer Szene durchführen. Diese wird allerdings hinsichtlich der Separation von Objekten in den meisten Fällen wesentlich ungenauer ausfallen als eine Oszillation, die sich über längere Zeit herausbilden kann. Da aber im Echtzeitbetrieb Eingangsbilder nicht unbegrenzt lange zur Verfügung stehen, ist die Möglichkeit einer solchen *One-Shot-Analyse* für natürliche wie für künstliche Segmentierungssysteme von Bedeutung.

Die Benutzung eines festen zeitlichen Gradienten mit retinozentrischem Ortsprofil über dem Sehfeld setzt allerdings voraus, daß eines der zu segmentierenden Objekte in der Szene bereits fixiert wird. Dies ist aber i.a. durch die Arbeit der Blicksteuerungsschicht gegeben. Das natürliche System zeigt sich an dieser Stelle wesentlich komplexer: Die Aufmerksamkeit (im Sinn einer beschleunigten Verarbeitung) kann offensichtlich auch ohne Veränderung der Blickrichtung verlagert werden (s. Kap. 2.5.2). Die Modellierung dieser speziellen Eigenschaft erscheint ungleich aufwendiger als der hier gewählte Weg, ohne daß *a priori* grundlegend neue funktionelle Eigenschaften zu erwarten sind, zumal beim

natürlichen Sehen die Blickrichtung der Aufmerksamkeit innerhalb von Sekundenbruchteilen nachfolgt. Deshalb wurde dieser Weg in der vorliegenden Arbeit nicht weiter verfolgt.

Um aufmerksamkeitsinduzierte Latenzen im Modell für die Segmentierung zu verwerten, genügt es, die visuelle Aufmerksamkeit als fest auf den aktuellen Fixpunkt gerichtet anzunehmen. Die zentralen Bereiche des Eingangsbildes werden dann gegenüber der Peripherie beschleunigt verarbeitet, so daß ein Objekt im Fixpunkt durch einen zeitlich eng zusammenhängenden Spikestrom repräsentiert wird, während periphere Objekte das Segmentierungsnetz entsprechend später (und eventuell dispergiert) erreichen.

Entsprechend dieser Festlegung wurden die in Kap. 5.2 beschriebenen Simulationen mit realen Szenen mit einem zweidimensionalen gaußförmigen, auf die Retina zentrierten zeitlichen Gradienten durchgeführt. Abb. 4.7 zeigt das Ergebnis bei der Anwendung auf die Separation zweier Rechtecke, von denen eines im Zentrum des Blickfeldes liegt. Da die zeitliche Dispersion – anders als im Beispiel aus Abb. 4.6 – nun kontinuierlich vom Ort abhängt, werden die beiden Objekte dem Segmentierungsnetz nicht in sich synchron präsentiert, sondern weisen auch eine innere Dispersion auf, die allerdings geringer ist als der zeitliche Abstand zwischen den Objekten. Δ bezeichnet hier die maximale Dispersion zwischen dem Zentrum und dem Rand des Gesichtsfeldes. Die Beschleunigung der Segmentierung bei wachsendem Δ ist aufgrund der schwächeren Vorseparation weniger ausgeprägt als im vorigen Beispiel, aber immer noch gut zu erkennen.

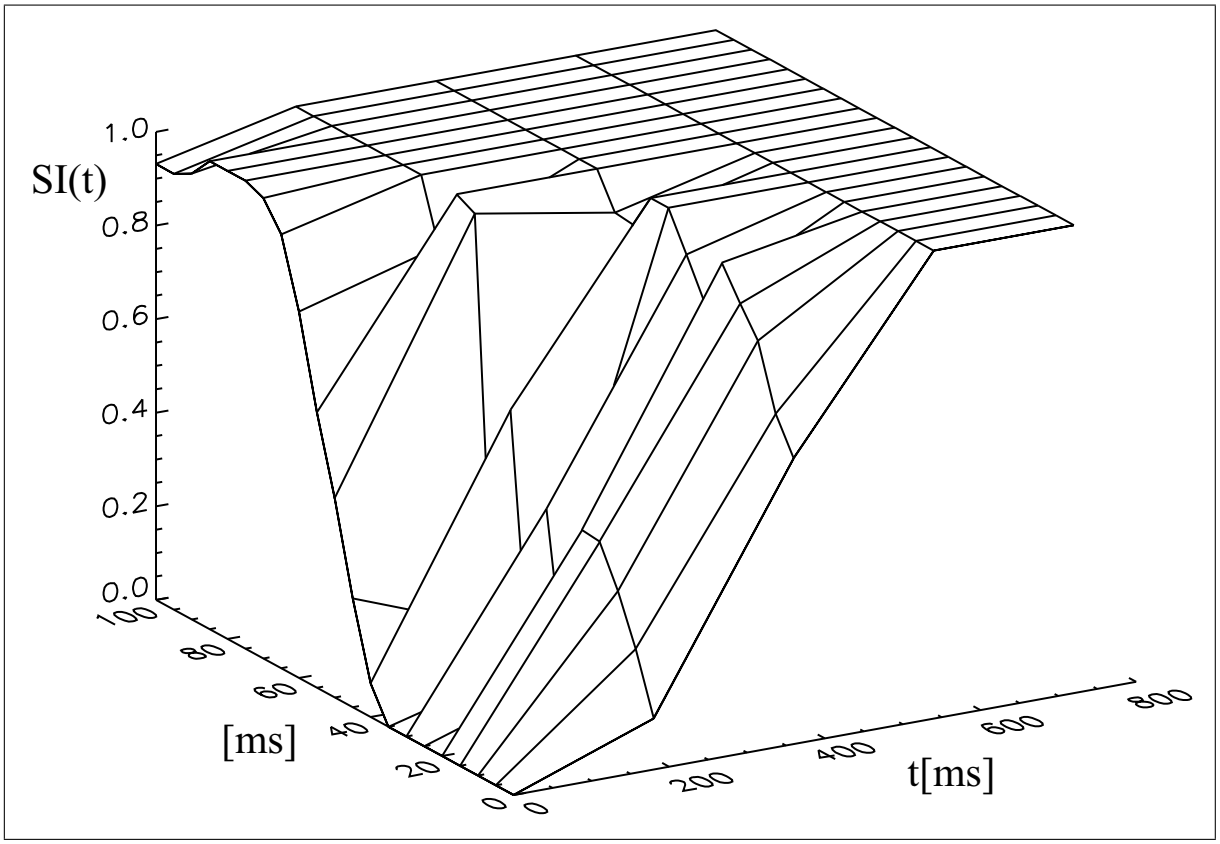


Abbildung 4.7: Separation zweier Rechtecke, die durch einen gaußförmigen Gradienten im Zeitbereich dispergiert werden. Das Zentrum des Gradienten liegt dabei auf einem der beiden Rechtecke, entsprechend einer Fixierung dieses Objekts.

5 Simulationsergebnisse mit realen Szenen

In diesem Kapitel soll die Anwendung der beschriebenen Netzwerke auf die Analyse bewegter, realer Szenen vorgestellt werden. Bei dieser Anwendung überlagern sich die Aufgaben von Objektverfolgung und Segmentierung: Eine komplexe Szene ist mit einer begrenzten Zahl von Neuronen nur dann sinnvoll zu bearbeiten (d.h. zu segmentieren), wenn die interessierenden Objekte als Blickziel ausgewählt und gegebenenfalls während ihrer Bewegung verfolgt werden. Diese Bewegung kann auch aus einer Eigenbewegung des Beobachters bzw. einer Überlagerung von Eigen- und Objektbewegung resultieren. Um die Analyse möglichst einfach und robust zu gestalten, muß die Segmentierung sich auf wenige Bildbereiche beschränken; das Segmentierungsnetz ist hier auf die 'Vorarbeit' der Aufmerksamkeitssteuerung angewiesen.

Zunächst werden die Verfolgungsergebnisse dargestellt (Kap. 5.1), danach die darauf beruhenden Segmentierungen (Kap. 5.2).

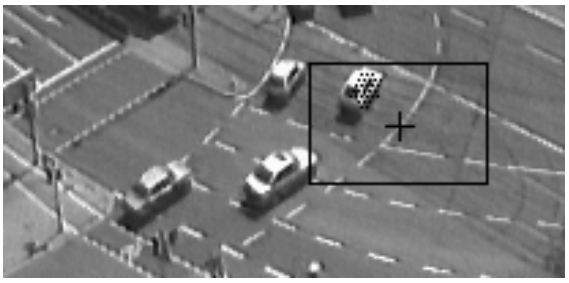
5.1 Verfolgungsergebnisse

5.1.1 Beispielszene 1: Durlacher Tor

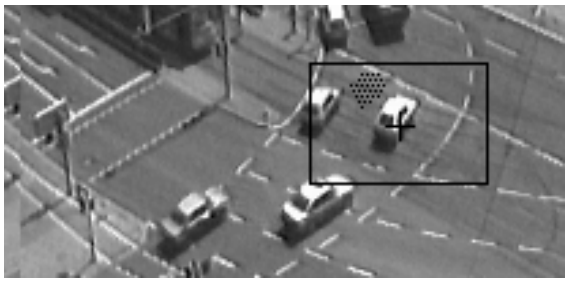
Abb. 5.1 zeigt eine Straßenkreuzung, die von oben mit feststehender Kamera aufgenommen wurde. Diese Szene eignet sich als Referenz für verschiedene Varianten der Verfolgung, weil die Autos als mögliche Blickziele klar definiert sind und sich sowohl durch Intensitätskontrast als auch durch ihre Bewegung vom feststehenden Hintergrund abheben. Schwierigkeiten entstehen durch die schlechte Auflösung; alle Autos sind unter 10×10 Pixel groß, so daß auch kleine Abweichungen das Target leicht aus dem Aufmerksamkeitsfokus (der selbst einen Durchmesser von ca. 10 Pixel hat) geraten lassen.

Zunächst wurde als einfachste Variante die Spikeaktivität der X-Zellen (aus zwei Auflösungsstufen mit Abtastweiten $d = 2$ und $d = 4$) direkt als Input auf das Aufmerksamkeitsnetz gegeben. Abb. 5.1 zeigt den bearbeiteten Bildausschnitt und die Entwicklung der Simulation. Am Anfang der Simulation steht das Target (Auto) in einer exzentrischen Position links oben relativ zur Blickrichtung. Die Aktivität in der Aufmerksamkeitsschicht konzentriert sich wie in Kap. 3.5 beschrieben schnell auf die Position des Autos (Abb. 5.1-56); nach 64 *bin* wird eine Sakkade zum Schwerpunkt der Aktivität ausgelöst. Abb. 5.1-65 zeigt die Situation unmittelbar nach der Sakkade. Durch die Verschiebung des retinozentrischen Koordinatensystems läßt sich die angezeigte Aktivität zunächst nicht sofort weiter sinnvoll auswerten (s. Kap. 3.5.7). Erst nach einer Suppressionszeit von 50 *bin* während

56

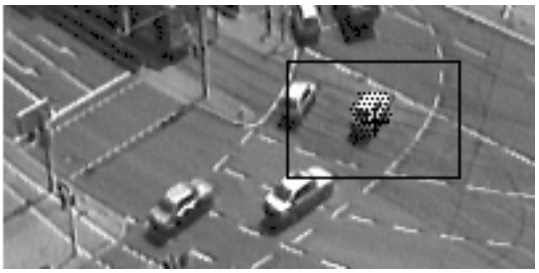


65



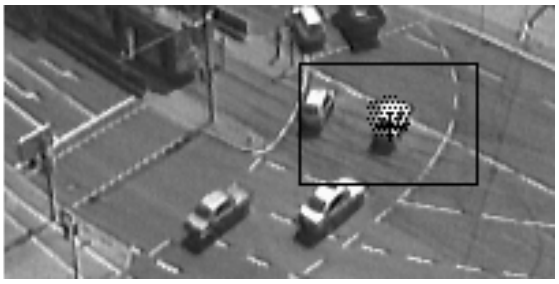
79

128

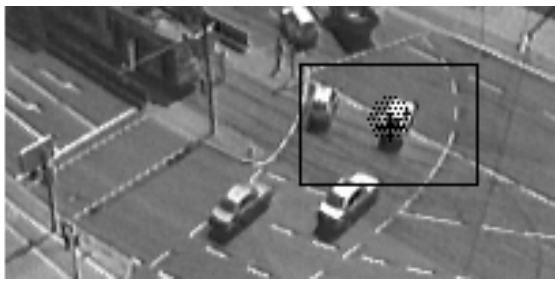


227

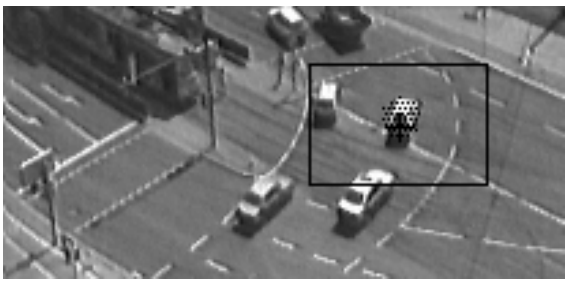
353



546

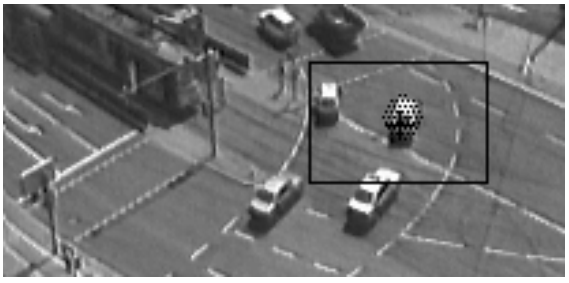


771



870

920



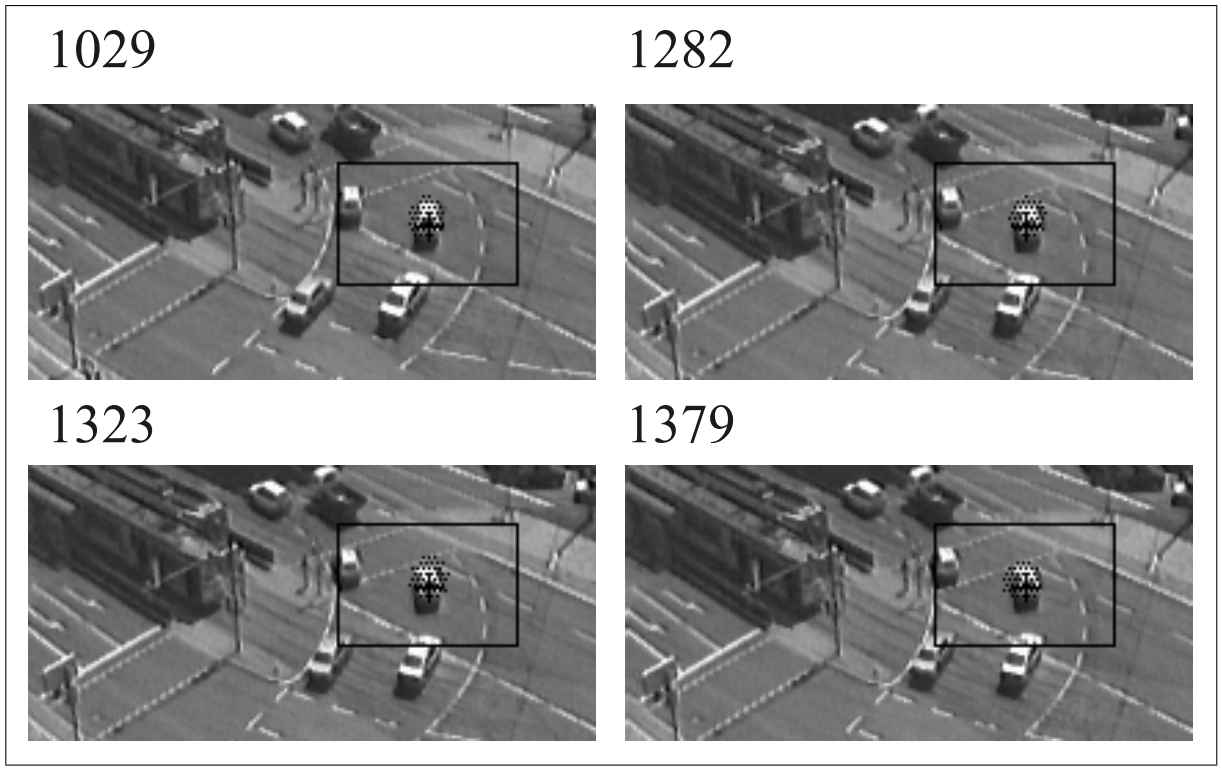


Abbildung 5.1: Simulationsbeispiel 1a: Durlacher Tor mit X-Zellen-Input. Der Rahmen im Bild zeigt den bearbeiteten Bildauschnitt. Das Kreuz in der Mitte des Rahmens gibt die momentane Blickrichtung an; die Einzelbilder sind durch die Nummern der dazugehörigen Zeitschritte markiert. **(56)** Am Anfang der Szene steht das zu verfolgende Auto außerhalb des Fixpunkts. Die Aktivität in der Aufmerksamkeitsschicht konzentriert sich auf die diese Position, so daß nach nach 64 bin eine Sakkade dorthin ausgelöst wird. **(65)** Unmittelbar nach der Sakkade müssen weitere Blickbewegungen und der Input zur Aufmerksamkeitsschicht zunächst unterdrückt werden **(79)**, so daß die gesamte Aktivität neu aufgebaut werden kann **(128)**. Man beachte, daß der gesamte Vorgang weniger als 0.1 *s* Echtzeit benötigt. Nach dem Neuaufbau der Aktivität wandert diese kontinuierlich mit dem Target mit und ermöglicht so die in Kap. 3.5.8 beschriebene glatte Folgebewegung **(227–1379)**.

der die Ausgangsaktivität der X-Zellen unterdrückt wird, baut sich die Aktivität im Zentrum der neuen Blickrichtung wieder auf (Abb. 5.1-128). In den weiteren Schnappschüssen der Simulation (Abb. 5.1-227–1379) ist zu sehen, wie die Aktivität kontinuierlich mit dem Auto mitwandert und so die in Kap. 3.5.8 beschriebene Folgebewegung herbeiführt. Ebenfalls zu erkennen ist der ständige leichte Schlupf, der notwendig ist, um die Folgebewegung in Gang zu halten.

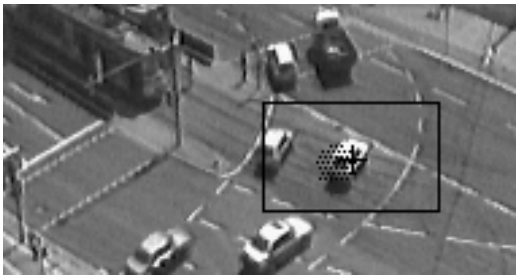
Abb. 5.2 zeigt die gleiche Szene, allerdings wird das vom Transientensystem erzeugte Merkmal *Bewegungskontrast* (s.Kap. 3.4) als Input für die Aufmerksamkeitsschicht verwendet. Der Verlauf der Simulation ist sehr ähnlich wie im vorhergehenden Beispiel (Intensitätskontrast als Input): Das auffälligste Blickziel im Bild wird mit einer initialen Sakkade in die Fovea gebracht; anschließend wird die Blickrichtung der Bewegung des Objekts nachgeführt.

Ähnliche Verfolgungen lassen sich für alle in der Szene vorhandenen bewegten Objekte (Autos) durchführen. Abb. 5.3 zeigt die dabei erzielten Trajektorien der Blickrichtung.

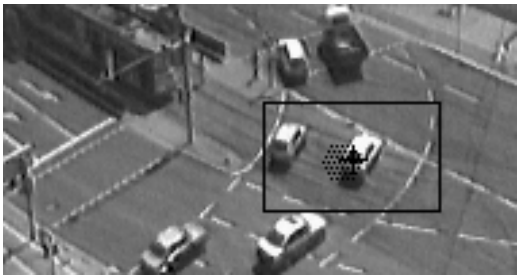
15

60

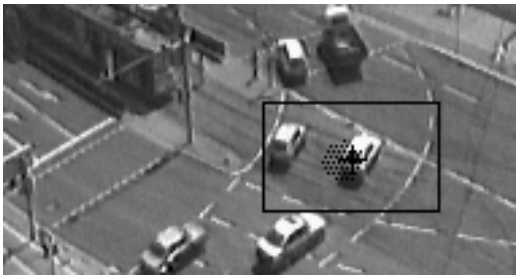
230



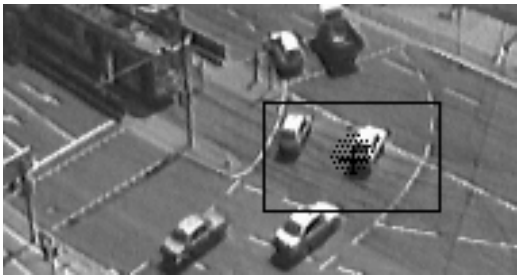
258



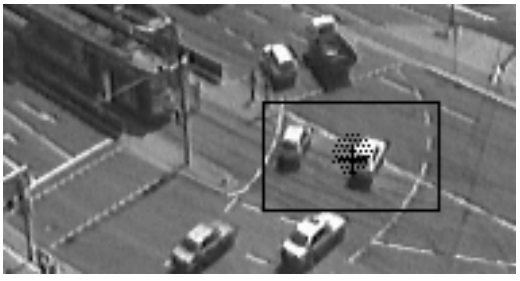
264



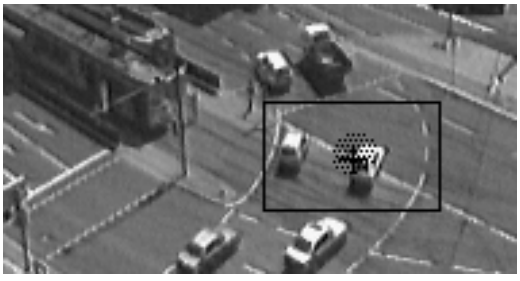
418



531



580



619

642

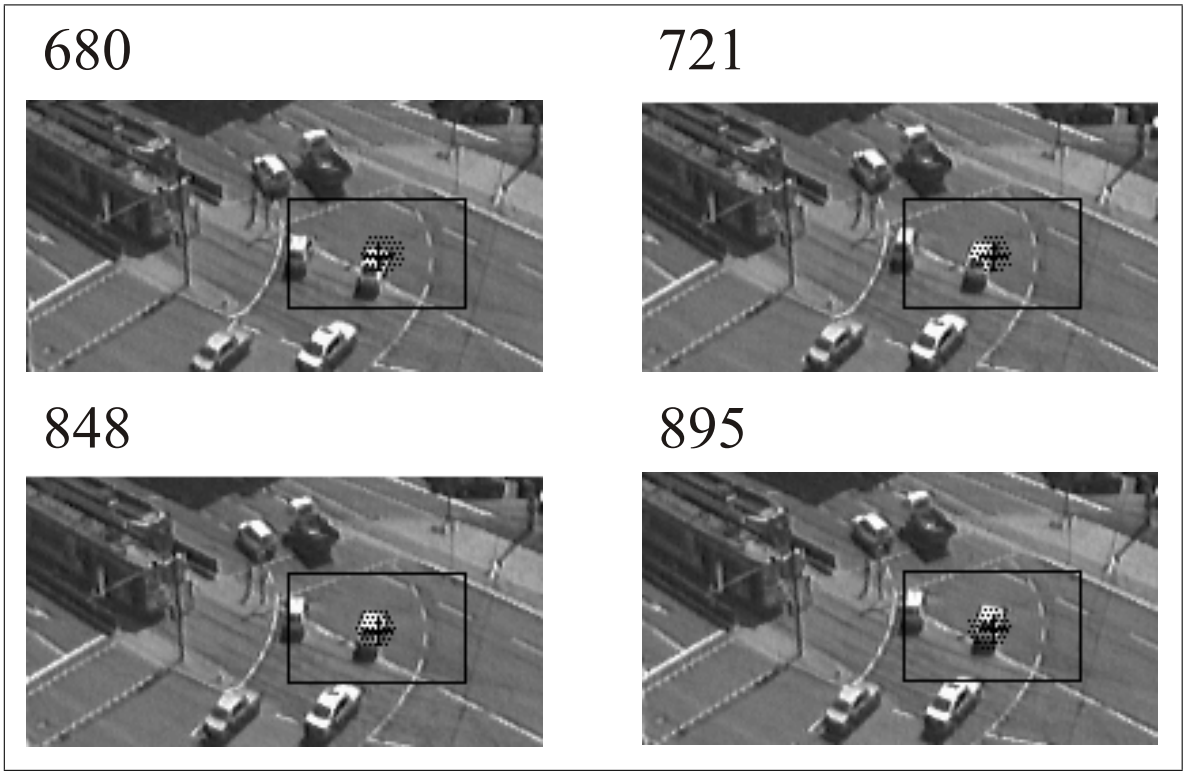


Abbildung 5.2: Simulationsbeispiel 1b: Durlacher Tor mit Aufmerksamkeits-Input vom Transientensystem. Der Verlauf der Simulation ist weitgehend der gleiche wie im vohergehenden Beispiel (die Nummern bezeichnen die zugehörigen Zeitschritte der Simulation): **(15)** Das zu verfolgende Auto steht außerhalb des Fixpunkts. **(60)** Die Aktivität in der Aufmerksamkeitsschicht konzentriert sich auf das neue Blickziel und löst aufgrund der peripheren Position des Blickziels eine Sakkade aus. Der Input wird für die Dauer der sakkadischen Suppression ΔT_{Sup} unterdrückt. **(230)** Nach der Sakkade erfolgt der Neuaufbau der Aktivität im Fixpunkt. **(258–895)** Die Blickrichtung wird der Bewegung des Autos kontinuierlich nachgeführt.

Abbildung 5.3: Übersicht über die Verfolgung vier verschiedener Autos im Beispiel Durlacher Tor. Die gleichen Trajektorien aus vier verschiedenen Simulationen sind in das Anfangsbild (links) und das Endbild (rechts) der Sequenz nach 1000 bin eingezeichnet. Das kurzzeitige 'Zurücklaufen' in der Mitte der linken oberen Trajektorie entstand beim Wechsel des Aufmerksamkeitsfokus von der rechten auf die linke Kante des verfolgten Autos.

5.1.2 Beispielszene 2: Fußgängerin

Abb. 5.5 zeigt die Verfolgung einer Fußgängerin. Als Input für die Aufmerksamkeitssteuerung wurde hier der Bewegungskontrast verwendet; in Abb. 5.4 ist der – über mehrere Zeitschritte aufintegrierte – Input separat dargestellt.

Da auch die Extraktion des Bewegungskontrasts letztlich auf der Aktivität der Detektoren für Intensitätskonstrast aufbaut, wandert der Schwerpunkt im Verlauf der Simulation vom Oberkörper der Person zu den Beinen. Wiederum macht sich hier die rein datengetriebene Arbeitsweise des Systems bemerkbar. Allerdings geht der Kontakt zur Person während der Simulation nicht verloren. Das Sichtfeld der Kamera wurde wie im vorhergehenden Beispiel relativ eng gewählt, da auf einer schnellen handelsüblichen UNIX-Workstation (DEC Alpha 21164 mit 500 MHz, 8 MB Second-Level-Cache, 256 MB RAM) bereits Kompilations- und Simulationszeiten in der Größenordung von Stunden anfielen.

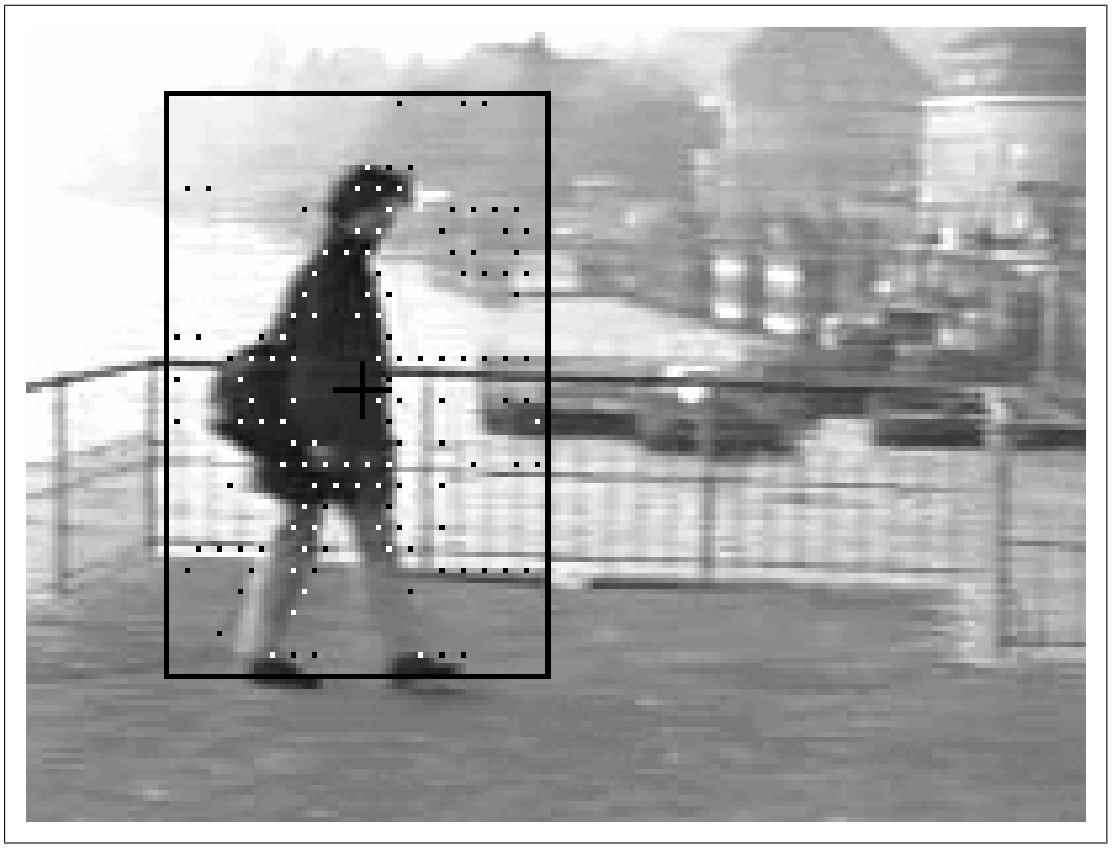


Abbildung 5.4: Aktivität der Bewegungskontrast-Detekoren bei der Szene aus Abb. 5.5 in einem Beispielframe ($t = 271bin$), aufintegriert über 8 Zeitschritte. Die Bewegungskontrast-Detektoren liefern den Input für die Aufmerksamkeitsschicht, deren Auswahldynamik dann das endgültige Blickziel bestimmt. Wie in Abb. 5.5 ist die Position der aktiven Neurone je nach Umgebung zur besseren Erkennbarkeit schwarz bzw. weiß markiert; inhaltlich besteht zwischen den beiden Markierungen kein Unterschied.

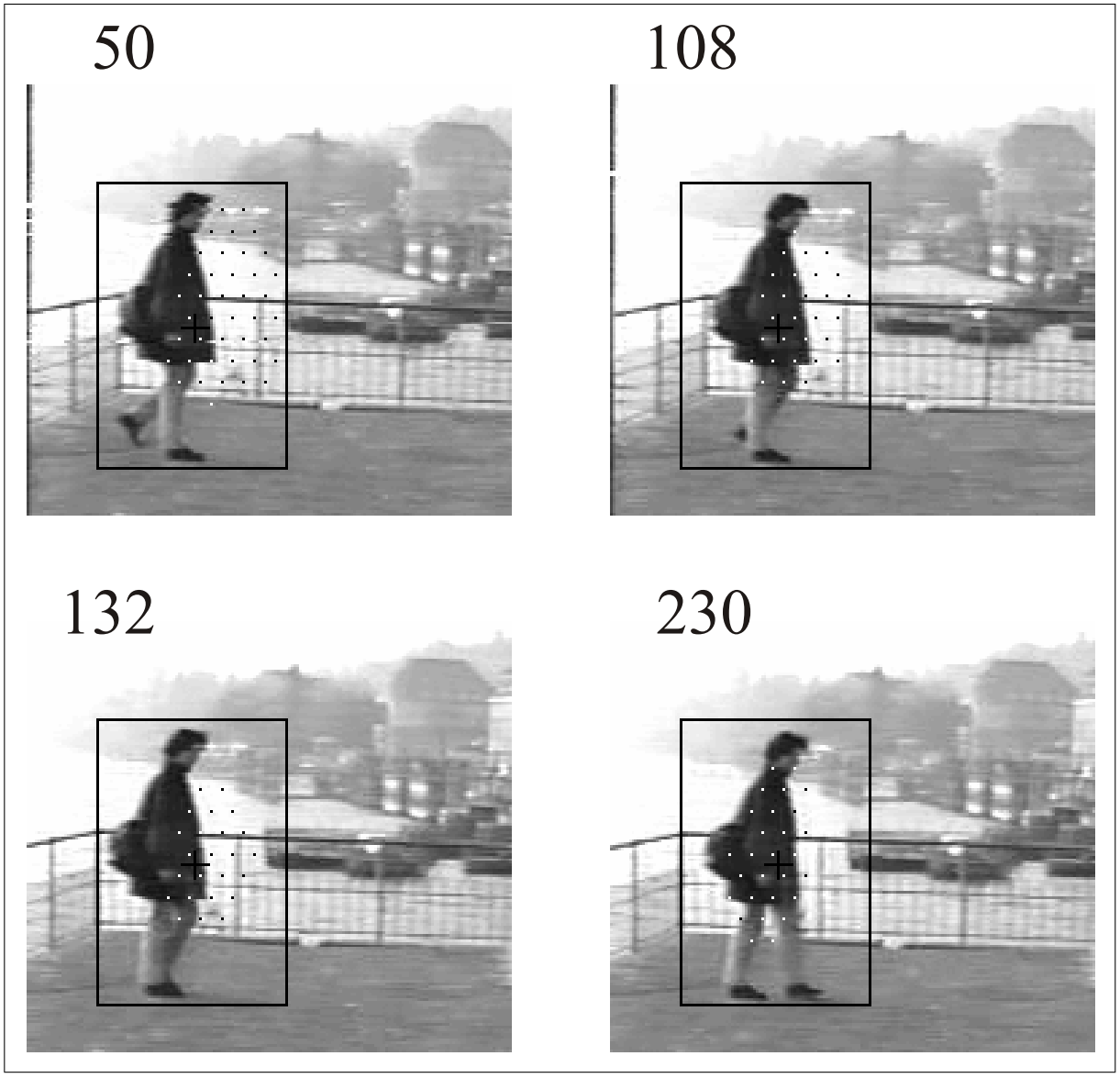
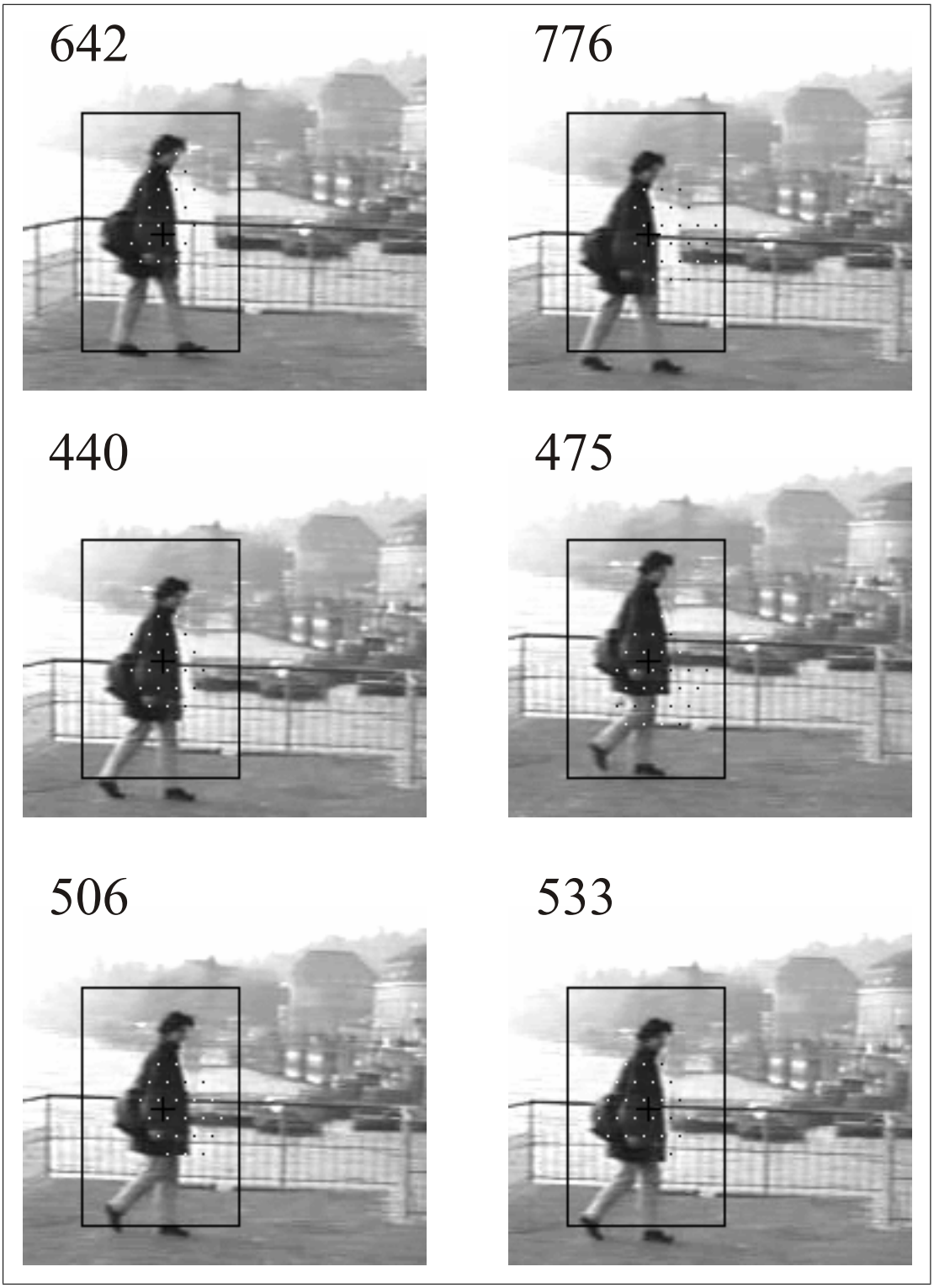


Abbildung 5.5: Simulationsbeispiel 2: Fußgänger (fortgesetzt auf den beiden folgenden Seiten). Die Zahlen bezeichnen wie in den vorhergehenden Beispielen die Zeitschritte, zu denen die Schnappschüsse festgehalten wurden. Die Aktivität der Aufmerksamkeitsschicht ist je nach Umgebungsfarbe schwarz bzw. weiß dargestellt, um eine Erkennung im Schwarzweißausdruck zu ermöglichen. Während der Simulation verlagert sich das Maximum der Bewegungskontrast-Detektion und damit das Input-Maximum für die Aufmerksamkeitschicht vom Oberkörper auf die Beine. Dies ist eine direkte Folge der rein datengetriebenen Arbeitsweise des Systems: Die Bewegungkonstrast-Detektion setzt auf die Aktivität der X-Zellen auf, die für Helligkeitskontrast empfindlich sind. Die zusätzliche Auswertung des Bewegungskontrastes bringt zwar eine starke Vorselektion des Inputs (vgl. Abb. 5.4), kann aber die Abhängigkeit von der retinalen Kontrastdetektion nicht völlig aufheben.


642
776
440
475
506
533

642

776

810

909

965

1032

5.1.3 Beispielszene 3a: Autobahn I

Abb. 5.6 zeigt einen Ausschnitt aus einer Autobahnfahrt bei normalen Sichtverhältnissen. Als Input für die Aufmerksamkeitsschicht wird ebenfalls wieder die Aktivität der Bewegungskontrastneuronen des Transientensystems verwendet. Anders als bei den bisher beschriebenen Simulationen steht hier die Kamera nicht fest, sondern führt eine Eigenbewegung aus, die sich im retinozentrischen System als Scheinbewegung der gesamten Szenerie äußert.

Zu Beginn der Szene (t=90 bin) ist der Blick noch nicht ausgerichtet. Nach kurzer Zeit (t=123 bin) baut sich in der Aufmerksamkeitsschicht Aktivität am Ort des vorausfahrenden Autos auf; dieses bindet den Fokus der Aufmerksamkeit. Anschließend erfolgt die entsprechende Sakkade; nach der sakkadischen Suppression ist noch Restaktivität des alten verschobenen Targets vorhanden, die mit dem neuen retinotopen Ort des Blickziels um den Fokus konkurriert (t=169 bin). Diese Konkurrenz fällt zunächst in die der Sakkade folgende Totzeit; das Auto übernimmt als neues Blickziel den Fokus für kurze Zeit wieder vollständig (nicht gezeigt).

Inzwischen ist allerdings ein weiteres potentielles Blickziel nähergerückt: Die scheinbare Geschwindigkeit der Ausfahrtbake am rechten Fahrbahnrand ist nach 231 *bin* so groß, daß zunehmend die in diesem Teil des Gesichtsfeldes angeordneten Geschwindigkeitsdetektoren und damit auch die für Geschwindigkeitskontrast empfindlichen Neurone aktiviert werden. Je genauer die retinotop gemessene Geschwindigkeit mit der 'optimalen Geschwindigkeit' der Detektoren übereinstimmt, desto stärker werden diese aktiviert. Nach 242 bin hat die Bake den Fokus vollständig vom Auto übernommen, was kurze Zeit später eine entsprechende Sakkade auslöst (t=257 bin).

Auch nach dieser Sakkade konkurriert eine gewisse Restaktivität (rechts oben im Bild) mit derjenigen am neuen retinotopen Ort des Blickziels; zusätzlich kommt kurzzeitig noch Input vom Fahrbahnrand rechts unten hinzu (t=331 bin). Für etwas weniger als 200 bin (ca. $\frac{1}{5}s$) bleibt die Bake als Blickziel erhalten; danach übernimmt das vorausfahrende Auto wieder den Fokus (t=457 bin). Nach der etwas ungenauen Rückkehrsakkade bei t=499 bin erfolgt bei 512 bin noch eine Korrektursakkade. Im weiteren Verlauf wird zwar nochmals Aktivität durch das Verkehrsschild und den im Hintergrund sichtbaren Höhenzug aufgebaut; zunächst fällt diese jedoch in die 50 bin dauernde Totzeit der Korrektursakkade. Nach dem Ende dieser Totzeit ist die Bake bereits so weit aus dem Gesichtsfeld gewandert, daß die periphere Aktivität zwar nicht vollständig unterdrückt wird (t=599 bin), andererseits aber auch nicht ausreicht, um eine neue Blickbewegung auszulösen. Im weiteren Verlauf bleibt der Blick auf das vorausfahrende Auto gerichtet.

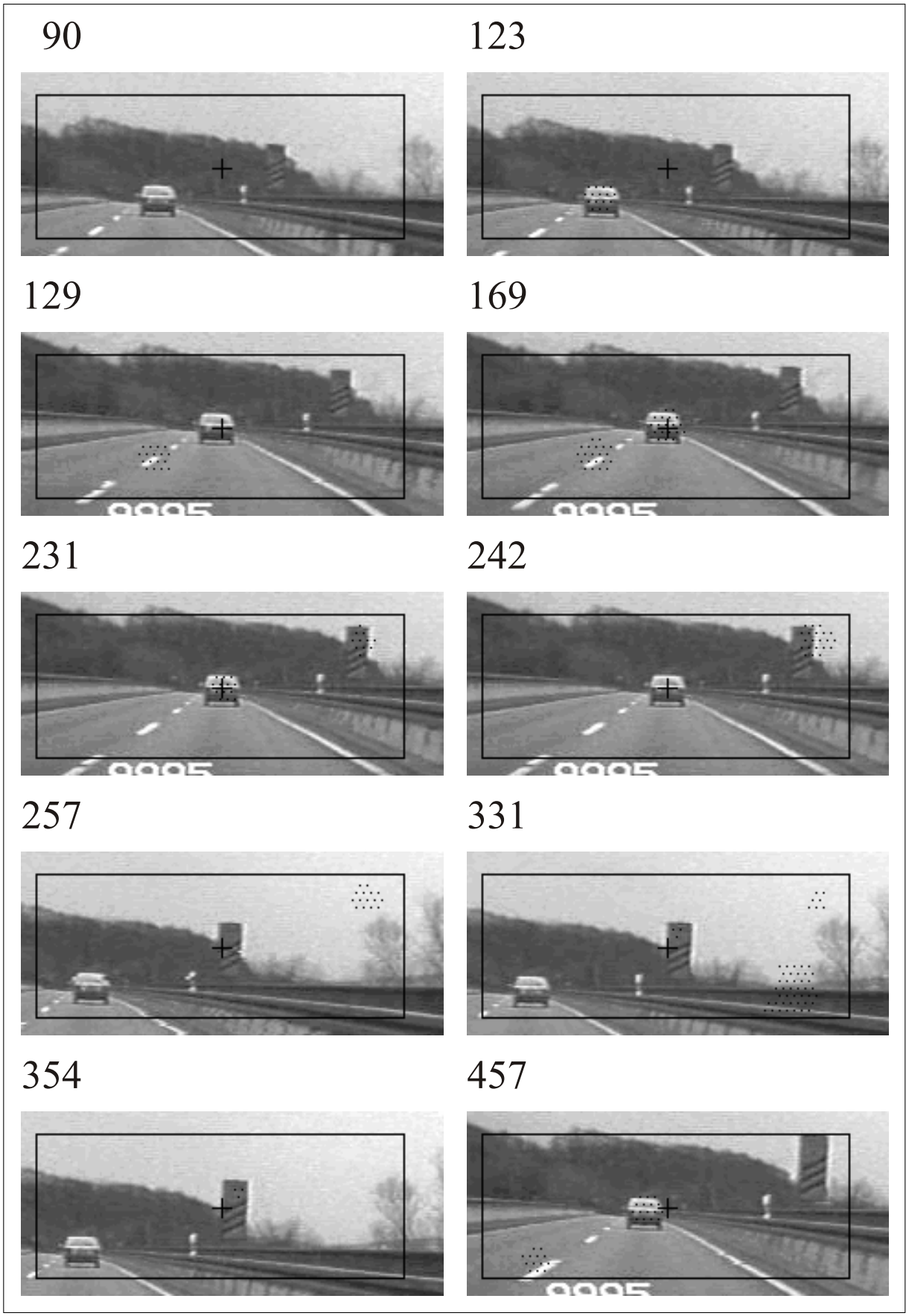
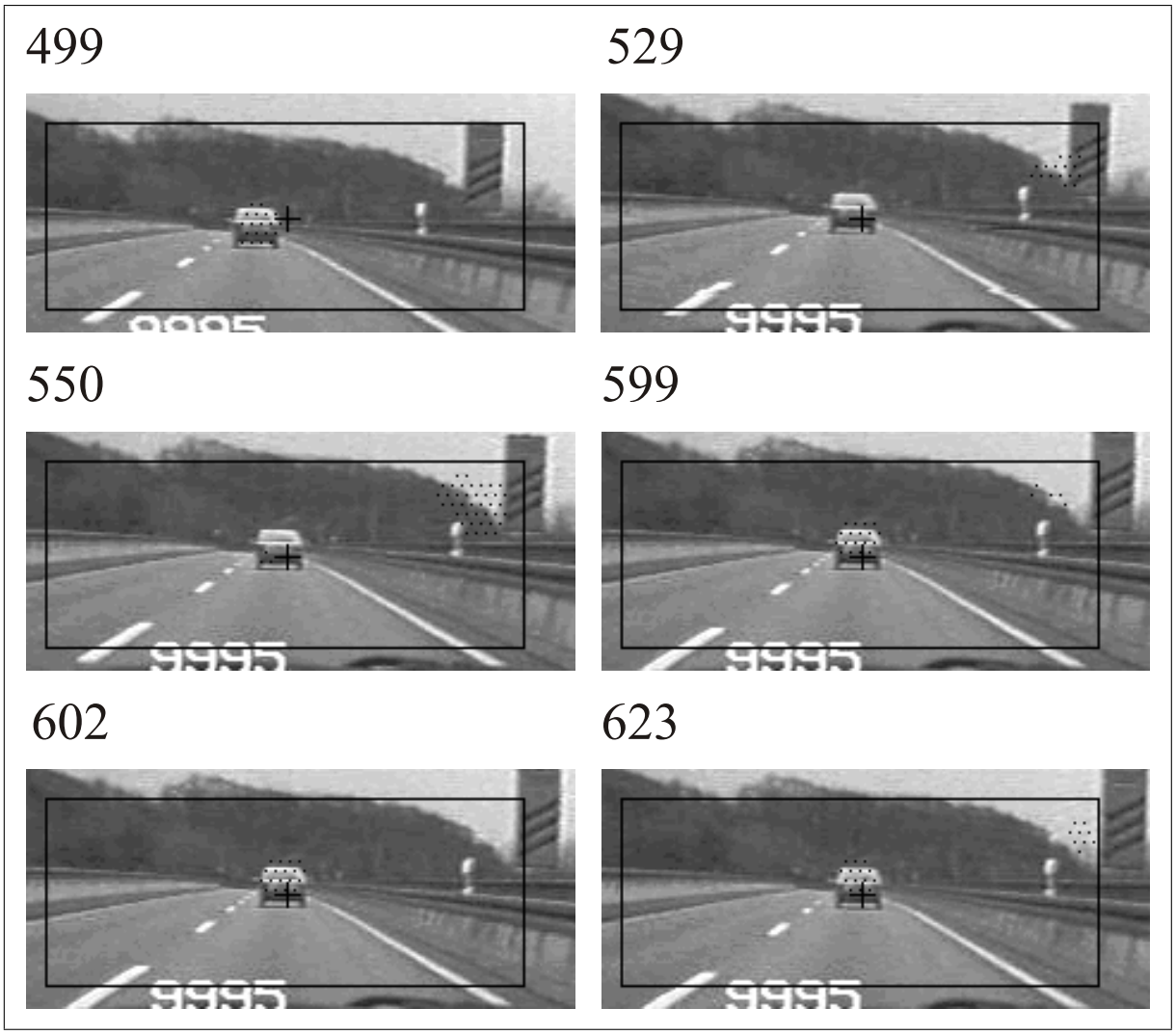


Abbildung 5.6: Simulationsbeispiel 3a: Autobahn I. Details s. Text.



5.1.4 Beispielszene 3b: Autobahn II

In der zuletzt beschriebenen Szene ist deutlich zu erkennen, daß mit dem vorausfahrenden Auto und der Ausfahrtbake zwei ungefähr gleich prominente Blickziele vorhanden sind. Dies äußert sich in der ständigen Konkurrenz der beiden Targets; keines kann auf Dauer die Aktivität des jeweils anderen vollständig unterdrücken; Inhibitionsdynamik in der Aufmerksamkeitsschicht befindet sich damit effektiv in einem bistabilen Bereich, in dem ein Wechsel von einem Target zum anderen bereits durch kleine Veränderungen in der Aktivität der Targets ausgelöst werden kann, vgl. dazu auch Kap. 3.5.9.

Abb. 5.7 zeigt die gleiche Szene, allerdings ist hier der Parameter h_{sel} etwas anders gewählt (0.68 statt 0.69). Der Beginn der Simulation verläuft ähnlich wie im vorhergehenden Beispiel: Zunächst ist das vorausfahrende Fahrzeug das stärkste potentielle Blickziel im Bild und wird somit nach einer initialen Sakkade bei t=128 bin fixiert. Nach etwa 200 bin beginnt sich am Ort des Verkehrsschilds Aktivität aufzubauen; bei 250 bin erfolgt eine Sakkade zum neuen Blickziel. Anders als in der vorher beschriebenen Simulation bleibt dieses jedoch stärkstes Target, wobei die Konkurrenz zwischen Auto und Schild

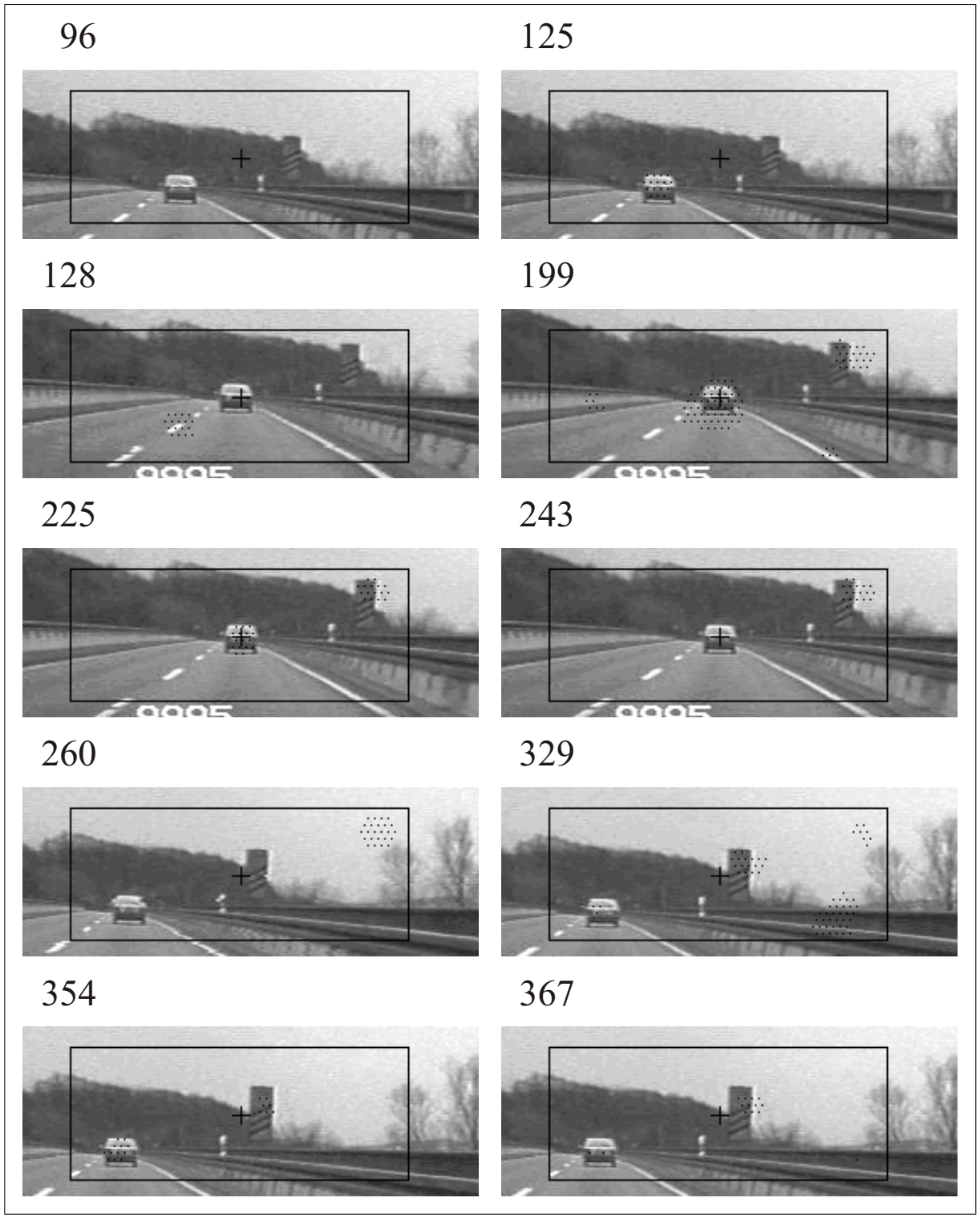
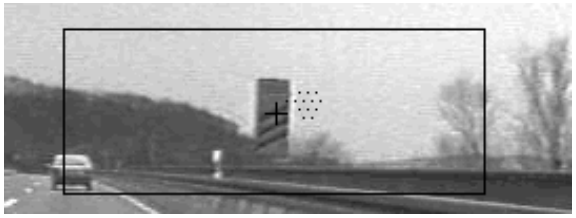
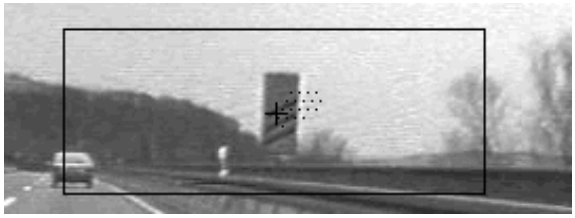


Abbildung 5.7: Simulationsbeispiel 3b: Autobahn II (Forts. nächste Seite)

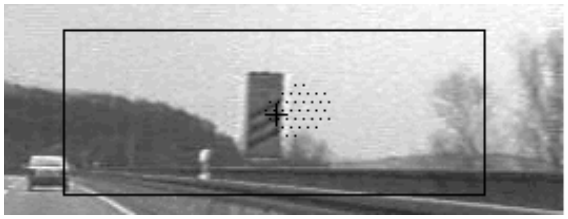
496



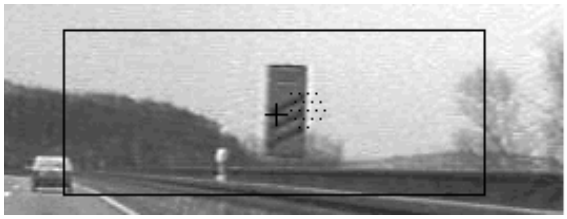
542



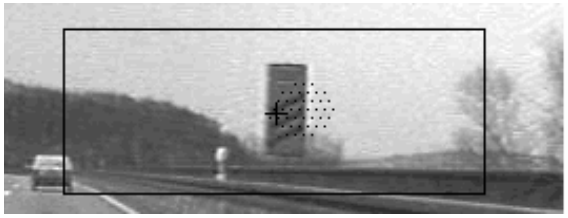
559



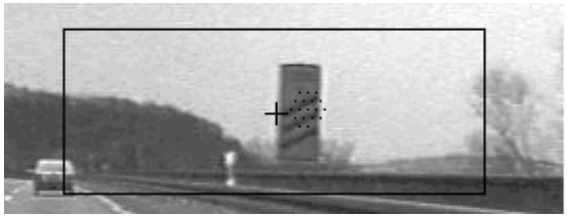
610



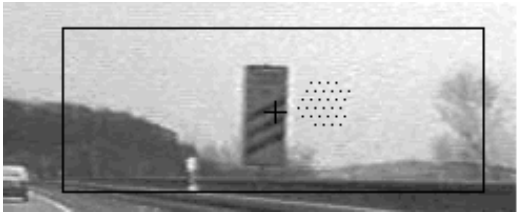
616



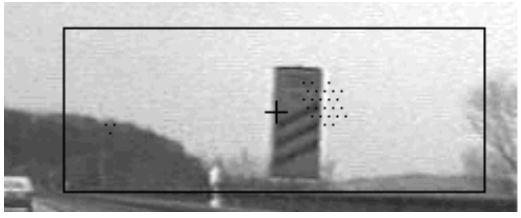
642



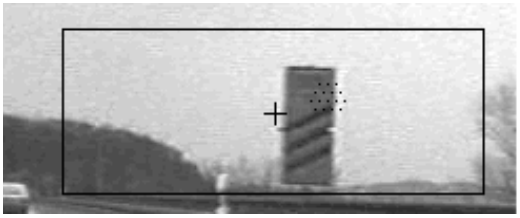
676



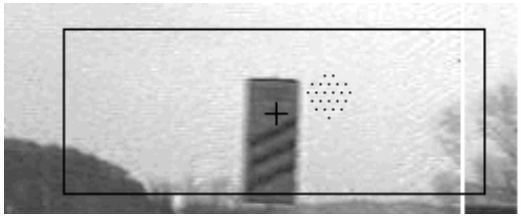
752



792



803



auch hier bestehen bleibt. Da allerdings die Blickrichtung der scheinbaren Bewegung des Schilds nach rechts ständig nachgeführt wird, wandert das Auto – in retinozentrischer Sichtweise – im gleichen Maß nach links, bis es schließlich nach ca. 600 bin aus dem Gesichtsfeld verschwindet. Da das System über keinerlei Objektgedächtnis verfügt, findet eine Rückkehrsakkade nicht statt; das Schild bleibt bis zum Ende der Simulation alleiniges Blickziel.

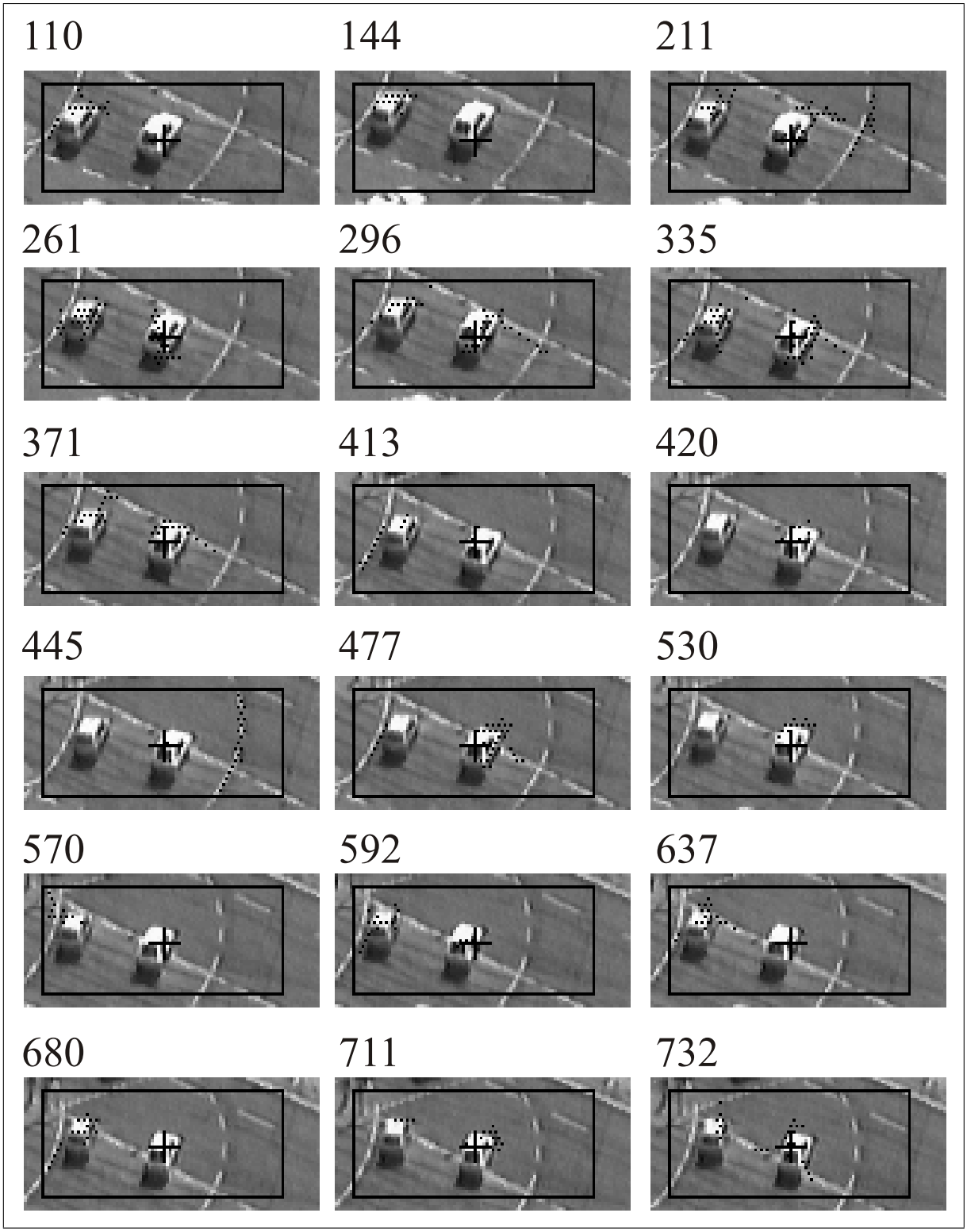
5.2 Segmentierungsergebnisse

Für eine ausführliche Darstellung der zeitlichen Segmentierung – auch komplexer – realer Szenen mit dem hier verwendeten Netz sei auf die Arbeit von Weitzel [1998b] verwiesen, die diesen Punkt zum zentralen Thema hat. Die vorliegende Arbeit untersucht lediglich die Besonderheiten, die sich durch die Einbeziehung der Aufmerksamkeitssteuerung bzw. Blickbewegungen ergeben:

- Die Bewegung der Kamera verändert den in das System eingespeisten Input signifikant: bei einer Folgebewegung wird das vorfolgte Objekt weitgehend ruhig auf der Retina gehalten, während der Hintergrund sich kontinuierlich bewegt. Je nach Objekt und Kamerabewegung können 'Verwacklungseffekte' auftreten.

- Die zeitliche Dispersion in der Vorverarbeitung liefert eine ständige Vorseparation der Szene, die – eine erfolgreiche Verfolgung vorausgesetzt – die Segmentierung robuster gegen die ständige Bewegung und Verwackelungseffekte machen kann.

Abb. 5.8 zeigt die Aktivität des Segmentierungsnetzwerkes bei Anwendung auf die Szene 'Durlacher Tor' (vgl. Beispiele im vorigen Abschnitt). Zur Erzeugung des Inputs wurde die Trajektorie der Verfolgung aus Beispiel 1b verwendet. Der Spikeoutput der X-Zellen, wurde auf das in Kap. 4 beschriebene, aus Kantendetektoren und Inhibitionsneuronen bestehende Segmentierungsnetz gegeben. Da in dieser Szene aufgrund der schlechten Bildauflösung ohnehin nur eine sehr beschränkte Extraktion der Objektkonturen möglich ist, wurden alle Kantenrichtungen einheitlich als schwarze Punkte dargestellt. Eine Zuordung zu den Objekten (linkes und rechtes Auto, Hintergrundelemente) ist aber auch in dieser vereinfachten Darstellung noch ohne weiteres möglich.


110
144
211
261
296
335
371
413
420
445
477
530
570
592
637
680
711
732

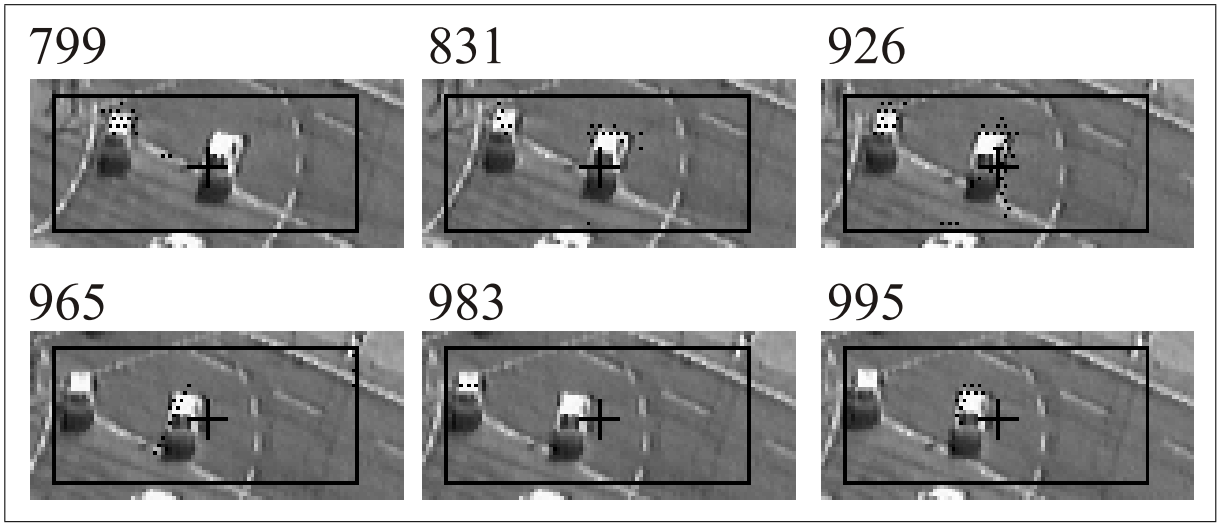
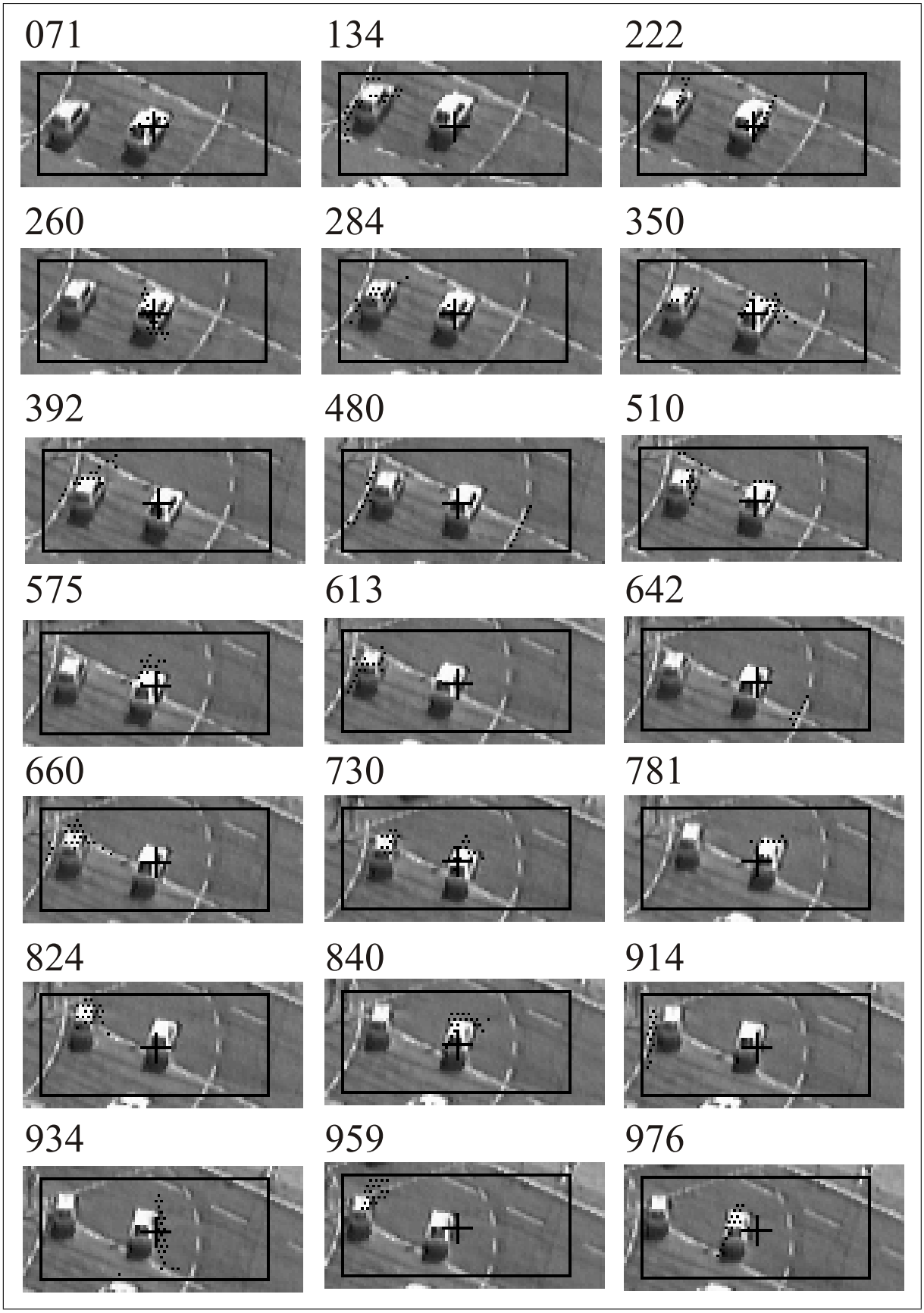


Abbildung 5.8: Simulationsbeispiel 4a: Aktivität des Segmentierungsnetzes bei Zugrundelegung der Kameratrajektorie aus Simulationsbeispiel 1b *ohne zeitliche Dispersion*. Die Aktivität der Kantendetektoren zeigt zwar den in Kap. 4.1 beschriebenen typischen oszillierenden Rhythmus; dargestellt ist jeweils das Aktivitätsmaximum eines Oszillationszyklus, wobei die Aktivität wie in den anderen Verfolgungsdarstellungen jeweils über 8 Zeitschritte aufintegriert wurde. Eine Phasentrennung der beiden Autos bzw. des Hintergrundes gelingt am Anfang der Simulation sowie in den Zeitschritten 413–680.

Abbildung 5.9: Simulationsbeispiel 4b (nächste Seite): Aktivität des Segmentierungsnetzes bei Zugrundelegung der Kameratrajektorie aus Simulationsbeispiel 1b *mit einem gaußförmigen zeitlichen Dispersionsgradienten* (vgl. Kap. 4.4.2). Die maximale Dispersion Δ_{max} betrug 25 bin, die Breite der Gaußglocke in x- und y-Richtung jeweils 15 Pixel. Die Separation gelingt bereits am Anfang der Sequenz und bleibt über den größten Teil der Simulation erhalten; lediglich in den beiden Oszillationszyklen um die Zeitschritte 350 sowie 730 geht sie teilweise verloren.


071
134
222
260
284
350
392
480
510
575
613
642
660
730
781
824
840
914
934
959
976

6 Konturdetektion in gestörten Bildern

Die Störungen, die in der Bildverarbeitung zu berücksichtigen sind, können verschiedener Natur sein. Es gibt sowohl 'intrinsische' (d.h. innerhalb des Bildverarbeitungssystems verursachte) Störungen als auch 'extrinsische' Störungen, deren Ursache in der äußeren Umwelt liegt. Die intrinsischen Störungen entstehen häufig aufgrund der mikroskopischen Arbeitsweise des jeweiligen Systems; typische Beispiele dafür sind das Kanalrauschen von Zellmembranen oder das thermische Rauschen von Halbleiterbauteilen. Extrinsische Störungen können z.B. durch elektromagnetische Einkopplung oder aufgrund optischer Behinderung des Lichtwegs, etwa durch Nebel, zustandekommen. Letzerer Fall ist einerseits für den praktischen Einsatz von Bildverarbeitungssystemen von großer Bedeutung, andererseits ist er nicht einfach zu formalisieren, da eine äquivalente ungestörte Situation bei realen Szenen immer hypothetisch ist.[1]

Um trotz dieser methodischen Schwierigkeiten, eine quantitative Behandlung des Problemfelds zu ermöglichen, wird in diesem Kapitel ein stochastisches Modell entwickelt, das es insbesondere erlaubt, die Auswirkungen der unterschiedlichen Typen von lateraler Kopplung wie sie in Kap. 3.3.5 vorgestellt wurden, zu analysieren. Die Liniendetektoren werden dabei **nicht** als Marburger Modellneurone mit innerer Dynamik behandelt, sondern als McCullough-Pitts-Neurone dargestellt. Das Eingangsbild wird also lediglich mit dem RF des jeweiligen Liniendetektors gefaltet und mit einer festen Schwelle Θ verglichen. Hinzu kommt, je nach betrachteter Variante, ein additiver oder multiplikativer Kopplungsterm von den Nachbardetektoren. Dieser verschiebt das Membranpotential in charakteristischer Weise und verändert so die Wahrscheinlichkeit einer überschwelligen Antwort.

Mit dieser vereinfachten Darstellung ist nun ein quantitativer Vergleich der beiden Kopplungsvarianten möglich; als Gütemaß für die Detektion fungiert dabei die Fehlerwahrscheinlichkeit im Sinne der klassichen Signalentdeckungstheorie, also die Wahrscheinlichkeit, daß eine Detektorantwort aufgrund einer Störung von 0 nach 1 wechselt oder umgekehrt.

Im folgenden wird zunächst das vereinfachte stochastische Modell im Detail erläutert. Auf dessen Grundlage werden die Auswirkungen der verschiedenen Kopplungstypen quan-

[1] Selbst wenn man ein Gebäude vom gleichen Blickwinkel aus an einem trüben und einem klaren Tag fotografiert, muß für einen quantitativen Vergleich der beiden Bilder bzw. ihrer Auswertung immer noch eine sorgfältige Kalibration, Normierung des Kontrastumfangs der Bilder etc. erfolgen. Selbstverständlich sind solche Untersuchungen unter geeigneten Laborbedingungen denkbar – den Rahmen der vorliegenden Arbeit würde dies aber bei weitem sprengen.

tifiziert; anschließend wird ihre praktische Leistungsfähigkeit anhand zweier künstlich erzeugter Bildstörungen bewertet. Abschließend wird die Frage diskutiert, inwieweit die vereinfachte Darstellung auch auf die kompliziertere Situation mit dynamischen Modellneuronen (und bewegten Eingangsbildern) anwendbar ist.

6.1 Das Neuron als Merkmalsdetektor: Statistische Formulierung

Neurone, die als lokale Merkmalsdetektoren arbeiten, haben die Aufgabe, die An- bzw. Abwesenheit eines Merkmals an einem bestimmten Ort im Eingangsbild möglichst zuverlässig zu detektieren. Anwesenheit des Merkmals wird im hier verwendeten Modell durch eine 1 am Ausgang angezeigt, Fehlen des Merkmals durch eine 0.

Statistisch betrachtet ist der Wert am Ausgang des Neurons eine Schätzung für das Vorhandensein des zu detektierenden Merkmals. Für eine Bewertung der Detektorleistung ist noch die Kenntnis des wahren Merkmalswerts nötig. Wir verwenden hier als Arbeitsdefinition für den wahren Merkmalswert die Antwort des Detektors bei völlig ungestörter Signalübertragung. Diese hängt selbstverständlich noch von den Parametereinstellungen der Neurone und ihren rezeptiven Feldern ab. Insbesondere muß die Schwelle Θ so eingestellt sein, daß weder alle vorkommenden Signale unterdrückt werden noch der Detektor spontan überschwellig ist. Beide Fälle sind für die Praxis ohne Bedeutung, da keine Information übertragen wird.

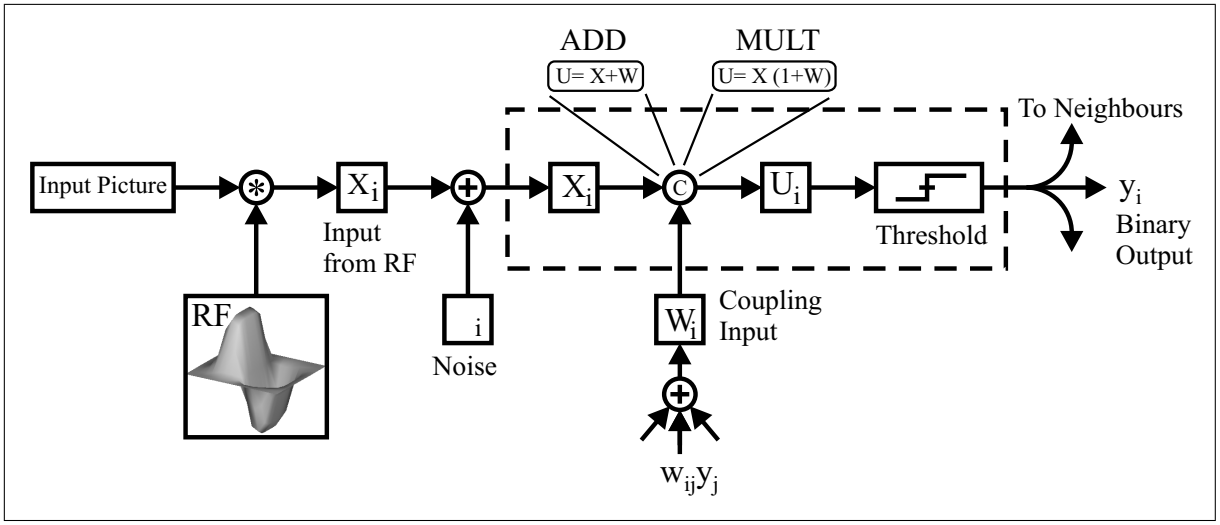


Abbildung 6.1: Stochastisches Modell für die Entstehung einer Detektorantwort bei verrauschter Konturdetektion. Die gestrichelte Linie deutet die Zellmembran des Neurons an.

Abb. 6.1 verdeutlicht die Entstehung des neuronalen Ausgangssignals in der Gegenwart von Rauschen. Zum deterministischen ungestörten Signal kommen zunächst Pixelrauschen (bzw. Rezeptorrauschen im biologischen System) und ggf. weitere Bildstörungen als stochastische Anteile hinzu. Für die Behandlung der neuronalen Antwort ist es aller-

dings günstiger, diese Einflüsse formal *nach* der Faltung mit dem RF des Neurons zu berücksichtigen, da ja nur das Ergebnis der Kantendetektion in die Berechnung des Membranpotentials eingeht. Das so gewonnene verrauschte Membranpotential wird dann noch mit dem additiven bzw. multiplikativen Kopplungsbeitrag der Nachbardetektoren zum Membranpotential verrechnet.[2]

Dieser Kopplungsbeitrag unterliegt als Linearkombination von Detektorantworten selbstverständlich denselben Rauscheinflüssen und müßte für eine exakte Behandlung ebenfalls als Zufallsgröße betrachtet werden. Wie die Korrelationsanalyse in Kap. 6.6 zeigt, muß man hier aber auf jeden Fall mit einer stochastischen Abhängigkeit der beiden Größen rechnen, die zudem von Szene zu Szene schwanken kann. Eine exakte Behandlung dieses Problems ist in der Literatur derzeit nicht bekannt und würde den Rahmen dieser Arbeit bei weitem sprengen.

Das bisher beschriebene Detektormodell erlaubt trotz der darin enthaltenen Näherungen eine quantitative Behandlung von Rausch- und Störeinflüssen sowie den Veränderungen, die sich durch verschiedene Typen von Nachbarschaftskopplung ergeben.

6.2 Rauschfreier und verrauschter Detektor

Wir betrachten zunächst eine eindimensionale Kette von Neuronen ohne Kopplung.

Für den Ausgang y_i des i-ten Neurons gilt im rauschfreien Fall:

$$\begin{aligned} y_i &= H(x_i - \Theta) \\ &= \begin{cases} 0 & : \quad x_i < \Theta \\ 1 & : \quad x_i \geq \Theta \end{cases} \end{aligned} \tag{6.1}$$

wobei x_i den kontinuierlichen Merkmalswert am Eingang des Detektors und H(x) die Heavisidesche Sprungfunktion bezeichnet.

Im verrauschten Fall tritt zu diesem 'wahren' Eingangswert ein Rauschterm φ_i als stochastische Komponente hinzu. Dieser Term beschreibt alle Rauscheinflüsse, die statistisch unabhängig von der Entstehung des ursprünglichen Eingangswertes x_i sind, also innere wie äußere Rauschphänomene.

Die Dichtefunktion der Zufallsvariablen φ_i sei $\psi(\varphi)$; die zugehörige Verteilungsfunktion wird mit $\Psi(\varphi)$ bezeichnet. In vielen Fällen (vor allem bei 'leitungstechnisch' begründetem Rauschen) wird $\psi(\varphi)$ in guter Näherung eine Normalverteilungsdichte sein. Bei natürlichen Störungen ist dies i.a. nicht der Fall. Diese treten häufig räumlich und zeitlich korreliert über mehrere Pixel hinweg auf und verschlechtern häufig die Kantendetektion, d.h. im Mittel sprechen weniger Detektoren als im ungestörten Fall an (z.B. verwaschene Konturen im Nebel). In diesem Fall hat $\psi(\varphi)$ einen negativen Erwartungswert.

[2] Da die Faltung mit dem RF des Neurons eine Linearkombination der einzelnen Pixel-Grauwerte darstellt, läßt sich die resultierende Wahrscheinlichkeitsverteilung für den Ausgangswert x_i der Kantendetektion berechnen, wenn die Rauschverteilung der Pixel nebst eventuellen stochastischen Abhängigkeiten bekannt ist. Im Fall eines normalverteilten Pixelrauschens ergibt sich wieder eine Normalverteilung, deren Breite kleiner oder gleich der ursprünglichen Rauschamplitude ist.

$$\Psi(\varphi) := \int_{-\infty}^{\varphi} \psi(x')\, dx' \tag{6.2}$$

Bezeichnen wir die so entstandene stochastische Eingangsvariable des Neurons mit X_i, so gilt für den Detektorausgang:

$$\begin{aligned} Y_i &= H(X_i - \Theta) \\ &= H(x_i + \varphi_i - \Theta) \end{aligned} \tag{6.3}$$

Damit wird Y_i selbst zu einer Zufallsvariablen mit den Ausprägungen 0 und 1. Die Stochastik dieser binären Variablen läßt sich vollständig durch die Wahrscheinlichkeit für $Y_i = 1$, d.h. die Wahrscheinlichkeit für einen überschwelligen Eingangswert darstellen:

$$\begin{aligned} P(Y_i = 1|x_i) &= P(X_i > \Theta) \\ &= \int_{-\infty}^{x_i - \Theta} \psi(x')\, dx' \\ &= \Psi(x - \Theta) \end{aligned} \tag{6.4}$$

Der Rauschterm hat somit den Effekt einer 'statistisch aufgeweichten' Schwelle, an die Stelle der Sprungfunktion tritt die Verteilungsfunktion des Rauschens.

6.3 Verrauschter Detektor mit Nachbarschaftskopplung

Die bisherige Argumentation ging von statistisch unabhängigen Neuronen aus, die als lokale Merkmalsdetektoren ohne jede Kopplung arbeiten. Um der Korrelationsstruktur natürlicher Bilder Rechnung zu tragen (und die Konturdetektion ggf. zu verbesseren), kann man durch eine exzitatorische Nachbarschaftskopplung die *bedingte Antwortwahrscheinlichkeit* für den Fall erhöhen, daß in der Umgebung bereits gleichartige Neurone geantwortet haben. Sinnvoll ist hier eine Kopplungstärke, die mit wachsendem Abstand abnimmt (beispielsweise exponentiell), da man zwischen weit voneinander entfernt liegenden Orten eine schwächere Korrelation als zwischen dicht benachbarten beobachtet (vgl. Kap. 6.6).

6.3.1 Additive Nachbarschaftskopplung

Realisiert man die Kopplung zwischen den benachbarten Neuronen mit gleich (bzw. ähnlich) orientierten rezeptiven Feldern durch additive Verbindungen, so kommt ein dritter Term W_i zum Membranpotential U_i hinzu, der den Einfluß der n angekoppelten, benachbarten Neurone widerspiegelt:

$$\begin{aligned} U_i &= X_i + W_i, \quad W_i = \sum_{j=1}^{n} w_{ij} Y_j \\ Y_i &= H(U_i - \Theta) \end{aligned} \tag{6.5}$$

wobei w_{ij} das Gewicht der Verbindung von Neuron j zu Neuron i bezeichnet.

Dieser Kopplungsterm W_i ist selbst wieder eine Zufallsvariable, da das Verhalten der Nachbarneurone (Index j) den gleichen Wahrscheinlichkeitsverteilungen unterliegt wie das betrachtete i-te Neuron. Für die folgenden Überlegungen zum Zusammenhang zwischen den Fehlern erster und zweiter Art spielt das genaue Zustandekommen von W_i allerdings keine Rolle. Als einzige Einschränkung betrachten wir eine rein exzitatorische Kopplung ($w_{ij} \geq 0$), so daß W_i nie negativ sein kann.

Durch die exzitatorische Nachbarschaftskopplung wird also das Membranpotential des Neurons i immer dann angehoben und somit die Wahrscheinlichkeit eines Ansprechens erhöht, wenn in der Nachbarschaft eines oder mehrere gleichartige Neurone bereits geantwortet haben. Anders ausgedrückt, wird die Wahrscheinlichkeit für einen *Fehler 1. Art* (nämlich eine eigentlich vorhandene Kante durch Rauschen zu verlieren) verringert.

Gleichzeitig wird aber auch die Wahrscheinlichkeit, einen *Fehler 2. Art* zu begehen, größer. Das bedeutet, daß die Kopplung auch dort Neurone überschwellig werden läßt, wo eigentlich keine Kante ist (und im fehlerfreien Fall auch nicht detektiert würde). Diese Gefahr setzt enge Grenzen für die Kopplungsstärken w_{ij}; diese müssen in ihrer mittleren Auswirkung auf jeden Fall unterhalb der Schwelle bleiben.

Das Problem wird besonders anschaulich, wenn man ein Linienende, beispielweise an einer Kontur-Ecke betrachtet. Hier würde ein Fehler 2. Art zu einem 'Ausfransen' beider Schenkel über den eigentlichen Kreuzungspunkt hinaus führen und z.B. eine nachgeschaltete Eckendetektion deutlich erschweren.

6.3.2 Multiplikative Nachbarschaftskopplung

Bei der multiplikativen Nachbarschaftskopplung von Eckhorn et al. [1990] (s. Kap. 3.3.5) werden die *additiven Feeding*-Verbindungen durch *multiplikative Linking*-Verbindungen ersetzt.

Der Wechselwirkungsterm wird bei dieser Art der Kopplung zusätzlich mit dem Eingangssignal des Detektors multipliziert, so daß sein Beitrag wiederum von dem ohne Kopplung erzeugten (verrauschten) Membranpotential X_i abhängt. Es ergibt sich folgende Situation:

$$\begin{aligned} U_i &= X_i \cdot (1 + W_i) \\ &= (x_i + \varphi_i) \cdot (1 + W_i) \\ &= (1 + W_i) \cdot x_i + (1 + W_i) \cdot \varphi_i \\ Y_i &= H(U_i - \Theta) \end{aligned} \tag{6.6}$$

Die Verteilung von X_i wird also wiederum nach rechts in Richtung höherer Antwortwahrscheinlichkeit verschoben und gleichzeitig verbreitert, da der Rauschterm um den

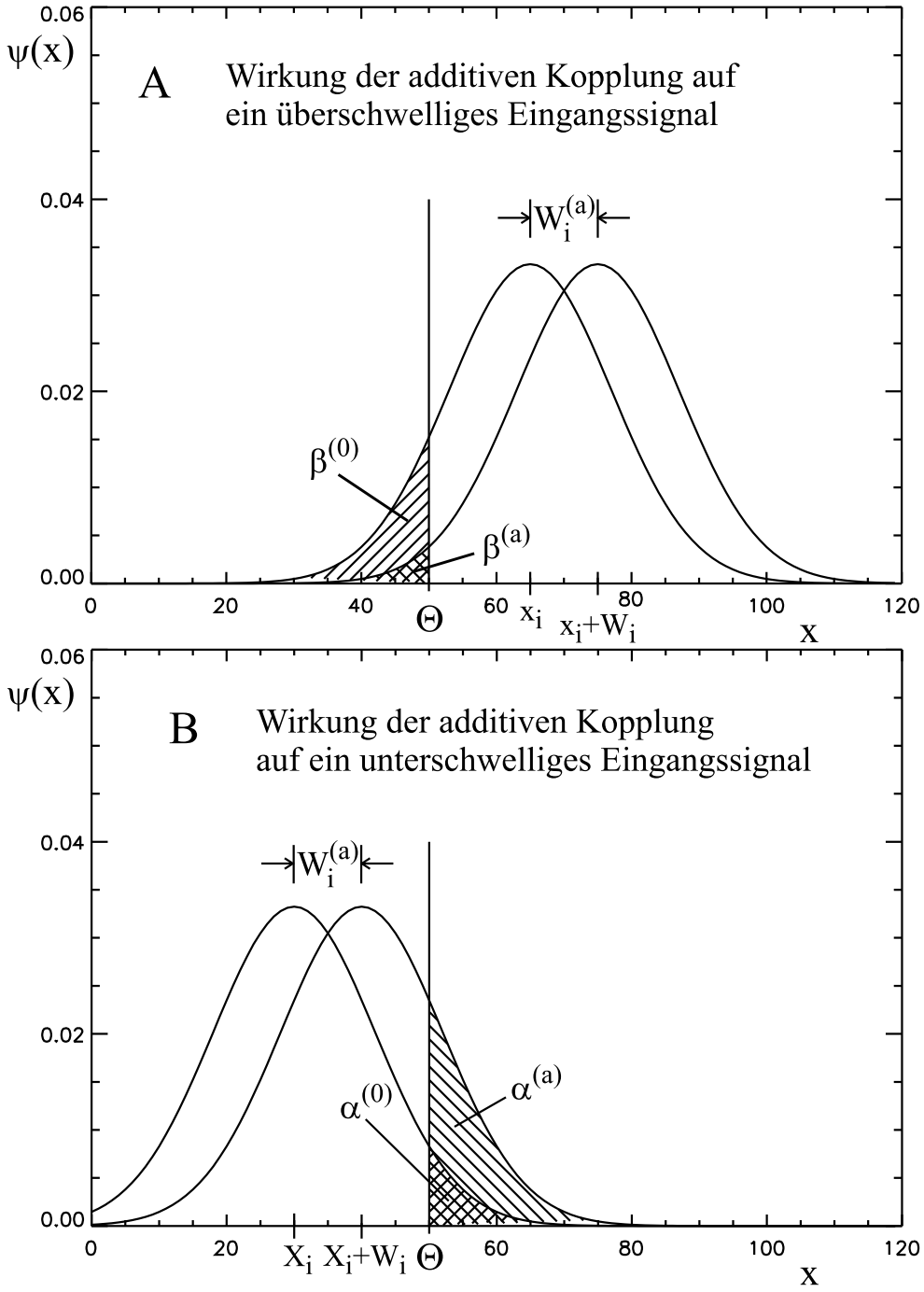


Abbildung 6.2: Einfluß der exzitatorischen, additiven Nachbarschaftskopplung auf die Wahrscheinlichkeitsverteilung des Membranpotentials für ein unter- und ein überschwelliges Eingangssignal. Die ursprüngliche Verteilung wird um den Betrag des Wechselwirkungsterms nach rechts verschoben. Dabei wird im überschwelligen Fall die Irrtumswahrscheinlichkeit β verringert, ihr Pendant α im unterschwelligen Fall dagegen vergrößert.

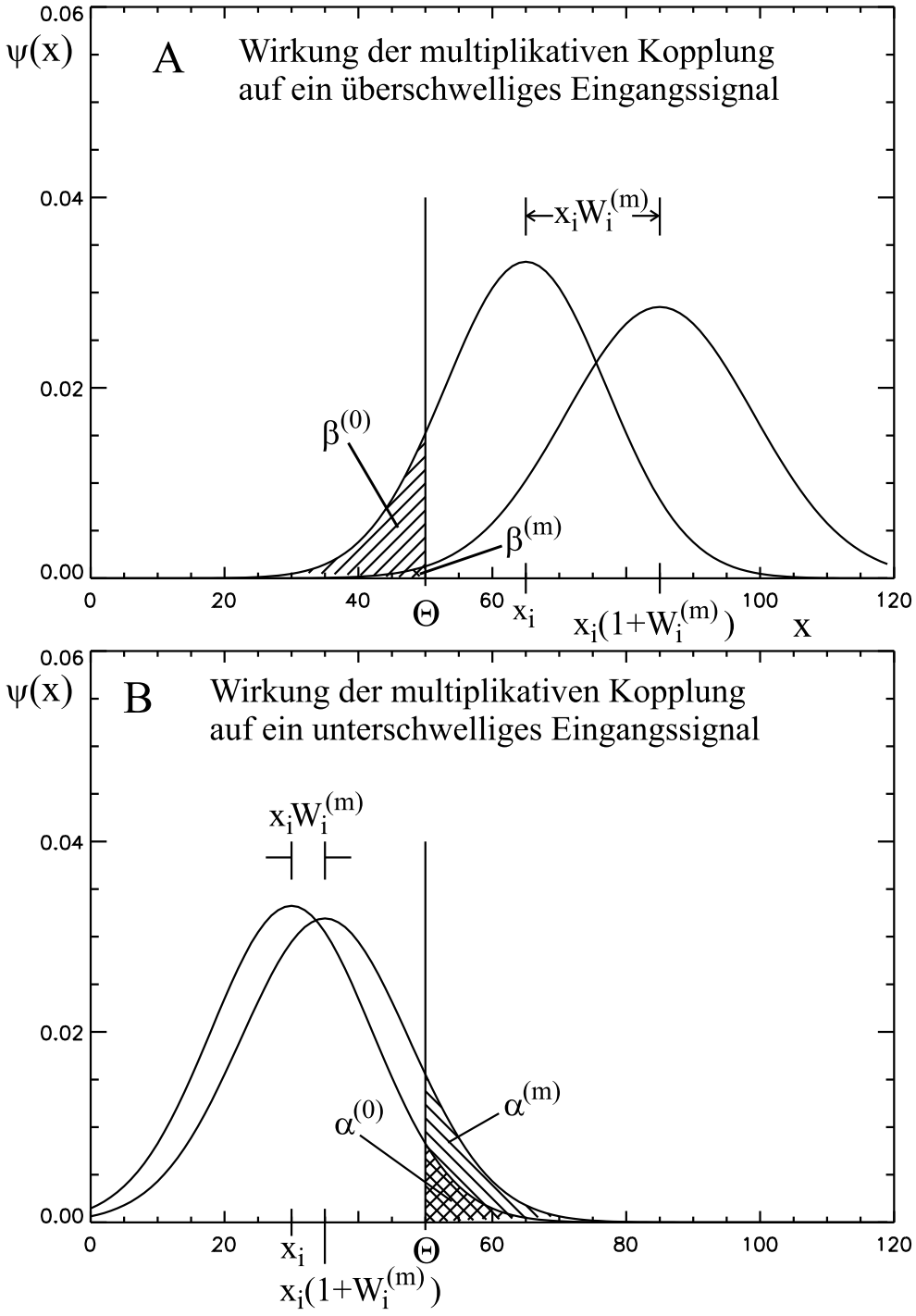


Abbildung 6.3: Einfluß der exzitatorischen, multiplikativen Nachbarschaftskopplung auf die Wahrscheinlichkeitsverteilung des Membranpotentials für ein unter- und ein überschwelliges Eingangssignal. Die ursprüngliche Verteilung wird um einen signalabhängigen Betrag nach rechts verschoben und gleichzeitig verbreitert. Wie bei der additiven Kopplung wird im überschwelligen Fall die Irrtumswahrscheinlichkeit β verringert, ihr Pendant α im unterschwelligen Fall dagegen vergrößert.

Faktor $(1+W_i)$ verstärkt wird. Dies ist eine unmittelbare Konsequenz der signalabhängigen Kennliniensteilheit bei gleichzeitiger Einwirkung von Feeding- und Linkingeinflüssen.

Abb. 6.2 und 6.3 veranschaulichen die Fehlerwahrscheinlichkeiten für die beiden Varianten der additiven und multiplikativen Kopplung.

6.4 Berechnung der Irrtumswahrscheinlichkeiten der Aktivierung für additive und multiplikative Nachbarschaftskopplung

Zwischen den beiden Irrtumswahrscheinlichkeiten für einen Fehler 1. bzw. 2. Art läßt sich für gegebene über- und unterschwellige Werte von x_i eine feste Beziehung angeben. Wir betrachten zunächst einen Spezialfall, in dem im Eingangssignal jeweils nur ein unter- sowie ein überschwelliger Wert von x_i vorkommen. Diese beiden Werte bezeichnen wir mit $\underline{x}_i$ (für unterschwellig) und $\overline{x}_i$ (für überschwellig), die Wahrscheinlichkeit für ihr Auftreten entsprechend als $p(\underline{x}_i)$ und $p(\overline{x}_i)$.

6.4.1 Additive Kopplung

Die Verteilung des verrauschten Signals entsteht bei additiver Nachbarschaftskopplung aus der Rauschverteilung $\psi(\varphi)$, die um den ungestörten Merkmalswert x_i sowie den Kopplungsbeitrag $W_i^{(a)}$ verschoben wird (wobei die Normierung erhalten bleibt):

$$s^{(a)}(X_i) = \psi(X_i - (x_i + W_i^{(a)})) \tag{6.7}$$

Die Irrtumswahrscheinlichkeit, ein überschwelliges Merkmal $\overline{x}_i$ fälschlich nicht zu detektieren, ist bei additiver Nachbarschaftskopplung:

$$\beta^{(a)} = \int_{-\infty}^{\Theta} \psi(x' - (\overline{x}_i + W_i^{(a)}))\, dx' = \Psi(\Theta - (\overline{x}_i + W_i^{(a)})) \tag{6.8}$$

Analog ist die komplementäre Irrtumswahrscheinlichkeit für unterschwelliges $\underline{x}_i$:

$$\alpha^{(a)} = \int_{\Theta}^{\infty} \psi(x' - (\underline{x}_i + W_i^{(a)}))\, dx' = 1 - \Psi(\Theta - (\underline{x}_i + W_i^{(a)})) \tag{6.9}$$

Zwischen den beiden Wahrscheinlichkeiten besteht über den Kopplungsbeitrag $W_i^{(a)}$ ein fester Zusammenhang. Löst man die Gleichung für $\beta^{(a)}$ nach $W_i^{(a)}$ auf, so ergibt sich:

$$W_i^{(a)} = \Theta - \overline{x}_i - \Psi^{-1}(\beta^{(a)}) \tag{6.10}$$

wobei Ψ^{-1} die Umkehrfunktion der Verteilungsfunktion bezeichnet.

Einsetzen in Gl. 6.9 liefert nach wenigen Umformungen:

$$\alpha^{(a)} = 1 - \Psi(\overline{x}_i - \underline{x}_i + \Psi^{-1}(\beta^{(a)})) \tag{6.11}$$

In den Zusammenhang der beiden Irrtumswahrscheinlichkeiten geht also bei gegebener Rauschverteilung nur die Differenz von über- und unterschwelligem Merkmalswert ein.

6.4.2 Multiplikative Kopplung

Wie oben beschrieben wird bei der multiplikativen Kopplung nicht nur das Signal, sondern auch der Rauschterm um den Faktor $f^{(m)} = 1 + W_i^{(m)}$ vergrößert. Die Wahrscheinlichkeitsverteilung des gestörten Signals ist gegeben durch:

$$s^{(m)}(X_i) = \frac{1}{f^{(m)}}\psi\left(\frac{X_i - x_i f^{(m)}}{f^{(m)}}\right) = \frac{1}{f^{(m)}}\psi\left(\frac{X_i}{f^{(m)}} - x_i\right) \tag{6.12}$$

Die Gleichungen für die Irrtumswahrscheinlichkeiten lauten im multiplikativen Fall:

$$\beta^{(m)} = \int_{-\infty}^{\Theta} s^{(m)}(X_i)\, dX_i = \Psi\left(\frac{\Theta}{f^{(m)}} - \overline{x}_i\right) \tag{6.13}$$

$$\alpha^{(m)} = \int_{\Theta}^{\infty} s^{(m)}(X_i)\, dX_i = 1 - \Psi\left(\frac{\Theta}{f^{(m)}} - \underline{x}_i\right) \tag{6.14}$$

Auflösen von Gl. 6.13 für β nach $\frac{\Theta}{f^{(m)}}$ ergibt:

$$\frac{\Theta}{f^{(m)}} = \Psi^{-1}(\beta^{(m)}) + \underline{x}_i \tag{6.15}$$

Damit ist der Zusammenhang zwischen $\alpha^{(m)}$ und $\beta^{(m)}$:

$$\alpha^{(m)} = 1 - \Psi(\overline{x}_i - \underline{x}_i + \Psi^{-1}(\beta^{(m)})) \tag{6.16}$$

Das Ergebnis ist identisch zum additiven Fall, kommt aber auf andere Weise zustande. Im multiplikativen Fall wird die Verteilung der verrauschten Merkmalswerte zwar auch nach rechts (in Richtung höherer x_i) verschoben, diese Verschiebung ist jedoch proportional zu x_i selbst. Daher könnte man erwarten, daß sich für unterschwellige x_i ein kleineres $\alpha^{(m)}$ als im additiven Fall ergibt und ebenso für überschwellige x_i ein kleineres $\beta^{(m)}$. Dieser Effekt wird jedoch von der Verbreiterung des Rauschens offensichtlich genau aufgehoben, so daß die Irrtumswahrscheinlichkeiten für beide Kopplungsvarianten identisch sind.

Da außer der Über- bzw. Unterschwelligkeit keine weiteren Annahmen über die betrachteten Werte von x_i in die Rechnung eingegangen sind, gilt das Ergebnis für jede beliebige Kombination eines über- und eines unterschwelligen Wertes von x_i. Somit ist es unabhängig von den Wahrscheinlichkeitsverteilungen, die das Auftreten der x_i bzw. das Zustandekommen von W_i beschreiben.

6.4.3 Antwortcharakteristik der Neurone mit Rauschen

Mit den Ergebnissen aus den beiden vorstehenden Abschnitten können wir nun die Auswirkung von Rauschen auf die in Kap. 3.3.5 angegebenen Antwortcharakteristiken zusammenfassend angeben.

Die deterministischen Kennflächen aus Abb. 3.13 werden durch Wahrscheinlichkeits-Kennflächen ersetzt, die die Antwortwahrscheinlichkeit eines Detektors bei gegebenem RF-Input und Kopplungsbeitrag angeben. Diese Darstellung zeigt Abb. 6.4. Daraus geht

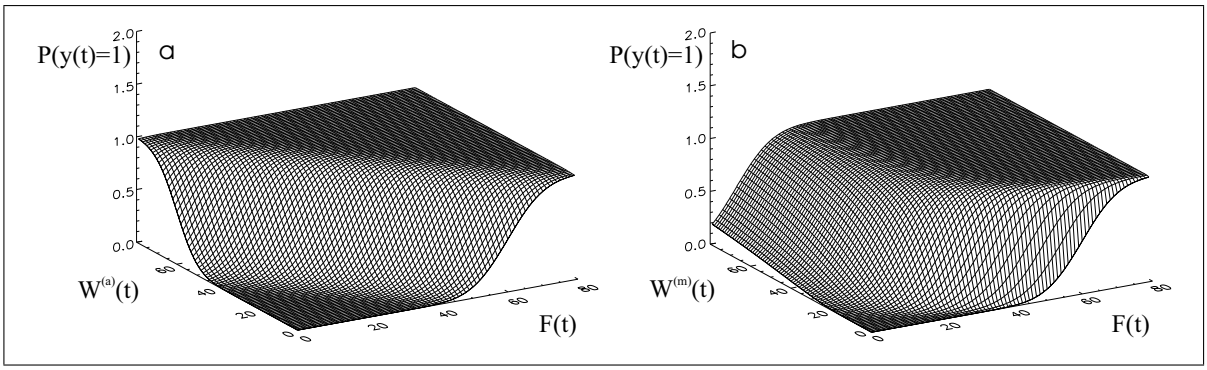


Abbildung 6.4: Ausgangscharakteristik des Modellneurons bei gleichzeitigem RF- und Kopplungsinput und Rauschen (Vgl. Abb. 3.14). Das (im Beispiel normalverteilte) Rauschen führt zu einer 'stochastisch aufgeweichten' Schwelle.

hervor, daß der Unterschied zwischen den Kopplungsvarianten mit zunehmendem Rauschen geringer wird. Insbesondere kann nun auch starker multiplikativer Kopplungseinfluß zusammen mit einer 'verrauschten Null' am Feeding-Eingang die Überschwelligkeit des Neurons herbeiführen.

Der in Abb. 3.3.5 angegebene Bereich II des Eingangsraums (der die Situation eines Neurons an einem Konturende repräsentiert), ist damit faktisch nach oben hin begrenzt; sobald die Rauschamplitude in der Größenordnung des Abstandes der begrenzenden Hyperbel zur Ordinate liegt, besteht auch bei der multiplikativen Kopplung die Gefahr, eine Kante über ihre eigentliche Ausdehnung hinaus in der neuronalen Repräsentation zu verlängern.

6.5 Die mittlere Irrtumswahrscheinlichkeit als Gütemaß für die Konturdetektion

Um die Leistungsfähigkeit der Konturdetektion zu quantifizieren, benötigen wir zunächst ein Gütemaß. Hierfür bietet sich die mittlere Irrtumswahrscheinlichkeit P_{Err} an, die durch folgenden allgemeinen Ausdruck gegeben ist:

$$P_{Err} = \int_{-\infty}^{\Theta} p(\underline{x})\alpha(\underline{x})\, d\underline{x} + \int_{\Theta}^{\infty} p(\overline{x})\beta(\overline{x})\, d\overline{x} \tag{6.17}$$

wobei für α und β je nach Art der Nachbarschaftskopplung Gl. 6.9 bis 6.13 einzusetzen sind. Die Bezeichnung der Integrationsvariablen als $\underline{x}$ bzw. $\overline{x}$ soll lediglich verdeutlichen, daß der erste Summand alle unterschwelligen Werte der Kantenstärke x umfaßt, während der zweite Summand alle bei ungestörter Detektion überschwelligen Werte beschreibt. Über die Ausdrücke für α und β gehen die Kopplungsgewichte w_{ij} als veränderliche Parameter ein.

Ziel der Optimierung ist es nun, P_{Err} bezüglich der w_{ij} zu minimieren. In der Praxis läßt sich dabei eine wesentliche Vereinfachung erzielen, wenn die grundsätzliche Form der Kopplungsmatrix feststeht (z.B. anisotrope, exponentiell abfallende Abhängigkeit vom Abstand), und die Optimierung bezüglich weniger Parameter durchzuführen ist.

6.6 Anwendungsbeispiel

6.6.1 Statistische Analyse der Eingangsbilder

Im folgenden wird die Auswirkung der Nachbarschaftskopplung an einem Beispiel mit zwei verschiedenen Arten von Rauschen verdeutlicht: pixelweise unabhängiges Gaußsches weißes Rauschen, wie es beispielsweise durch einen gestörten Übertragungskanal entsteht, und eine quasi-natürliche Störung, die in etwa der Sicht bei Nebel oder durch eine verschmutzte Windschutzscheibe entspricht. Als Referenz dient das ungestörte Bild (Abb. 6.5).

Die Untersuchung von Störprozessen in natürlichen Bildern gestaltet sich schwierig, da normalerweise keine ungestörte Referenz zur Verfügung steht. Für eine quantitative Analyse ist zudem eine gute Kalibrierung der Aufnahmen notwendig, was mit erheblichem Aufwand verbunden ist. Aus diesem Grund wurde die 'nebelartige' Störung künstlich durch lokale Glättung des Eingangsbildes mit einer Gaußmaske erzeugt, deren Breite von Ort zu Ort zufällig variiert. Dieser Prozess bildet in stark vereinfachter Form die Streuung von Licht an Wolken aus Partikeln oder Wassertröpfchen nach; auch hier entsteht durch die Überlagerung vieler zufälliger Einzelstreuungen eine räumliche Normalverteilung.

Diese Art der Bildveränderung hat den Vorteil, dass der Gesamtbetrag der Signalenergie im Bild erhalten bleibt, so daß der Arbeitsbereich der Neurone in etwa gleich bleiben kann. Die lokale Glättung des Bildes mit einer normierten Gaußmaske entspricht einer Tiefpaßfilterung, bei der Signalenergie von hohen zu niedrigen Ortsfrequenzen umverteilt wird. Da die verwendeten Kantendetektoren (s.u.) als hochfrequent eingestellte räumliche Bandpässe arbeiten, sinkt durch die Störung die mittlere Kantenstärke ab. Dies entspricht dem Wahrnehmungseindruck 'verwaschener' oder unscharfer Objektkonturen.

Die Eingangsbilder wurden zur Kantendetektion mit dem in Abb. 6.1 gezeigten RF in acht Orientierungen gefaltet; das Ergebnis der Faltung wird im folgenden als *Kantenstärke* bezeichnet. Um den Ausgangswert der lokalen Kantendetektoren zu bestimmen, wurde jeweils die lokale Kantenstärke mit der Schwelle Θ verglichen. Die Schwelle war so gewählt, daß zwischen 5% und 10% der Pixel überschwellig wurden. Dies entspricht exakt der Arbeitsweise einer retinotopen Karte von Kantendetektoren, von denen jeder als McCulloch-Pitts-Neuron arbeitet.

Für die weitere Analyse wurde zunächst für das ungestörte Eingangsbild die Häufigkeitsverteilung $p(x)$ der Kantenstärken bestimmt. Wie Abb. 6.7 zeigt, läßt diese sich recht gut durch eine Exponentialverteilung annähern, die in der Nähe des Nullpunkts in eine algebraische Verteilung übergeht. Das bestätigt die oben angestellte Vermutung, daß Bildorte mit einer Kantenstärke in der Nähe von Null die weit überwiegende Mehrheit stellen. Anders ausgedrückt, handelt es sich bei der Kantenstärke um ein *spärlich verteiltes* Merkmal.

Anschließend wurde die räumliche Korrelation der Kanten untereinander bestimmt, d.h. die räumliche Autokorrelation des Kantenbildes. Wendet man dieses Verfahren auf das binarisierte Kantenbild (also die Ausgangssignale der Detektorneurone) an, so entspricht die auf 1 normierte Autokorrelation genau der bedingten Wahrscheinlichkeit, in einem bestimmten zweidimensionalen Abstand von einem Neuron wieder eine gleichgerichtete Kante zu finden. Wie Abb. 6.6 zeigt, ist erwartungsgemäß eine starke Anisotropie

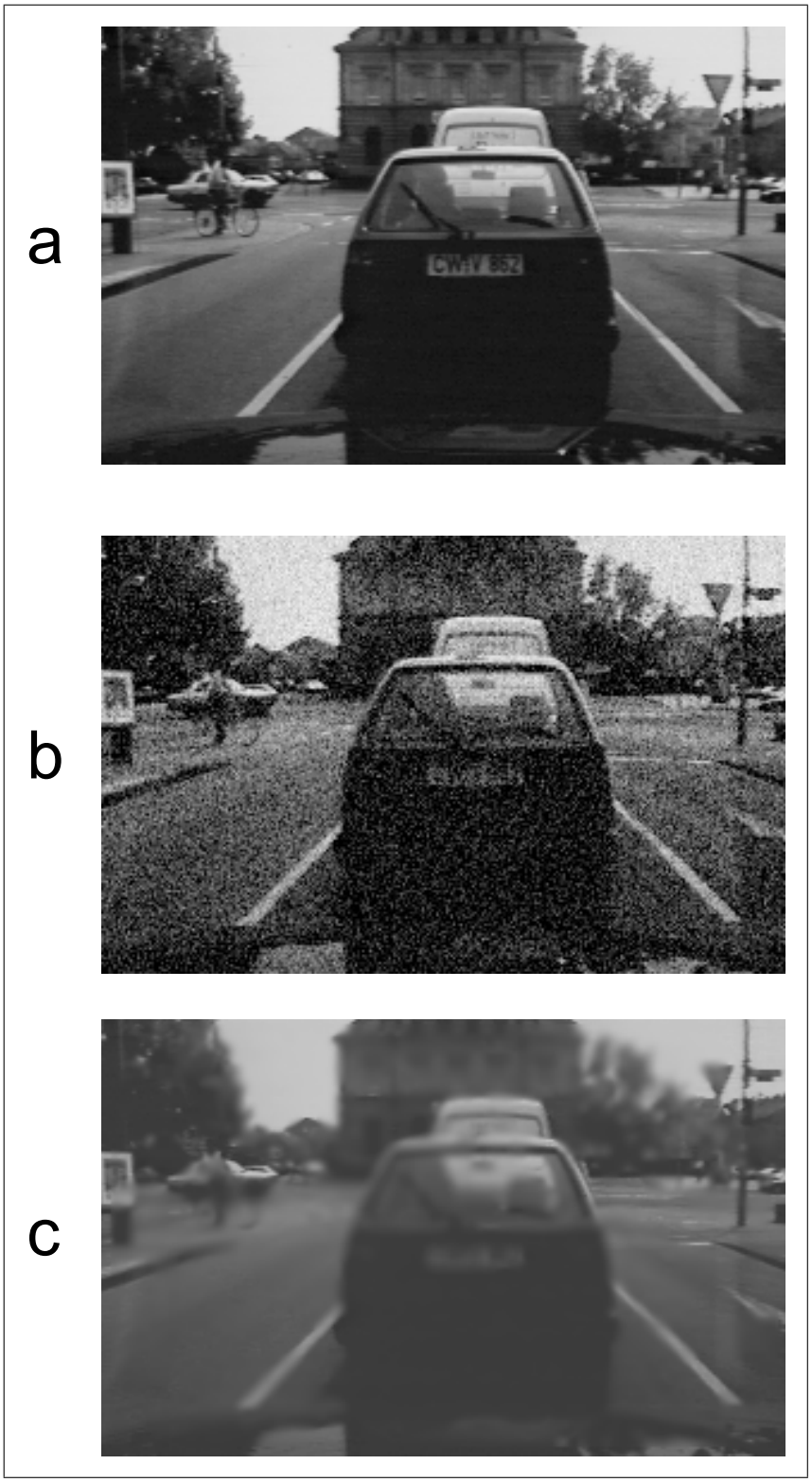


Abbildung 6.5: Beispiele für zwei unterschiedliche Rauscharten. **(a)** ungestörtes Eingangsbild **(b)** Gaußsches weißes Rauschen, das pixelweise unabhängig ist **(c)** Eine quasi-natürliche Störung, die den Effekt schlechter Sichtbedingungen nachbildet

in Iso-Orienterung der Kantenrichtung vorhanden. Der Gesamtverlauf der Funktion läßt sich recht gut durch eine zweidimensionale Exponentialfunktion annähern (nicht eingezeichnet).

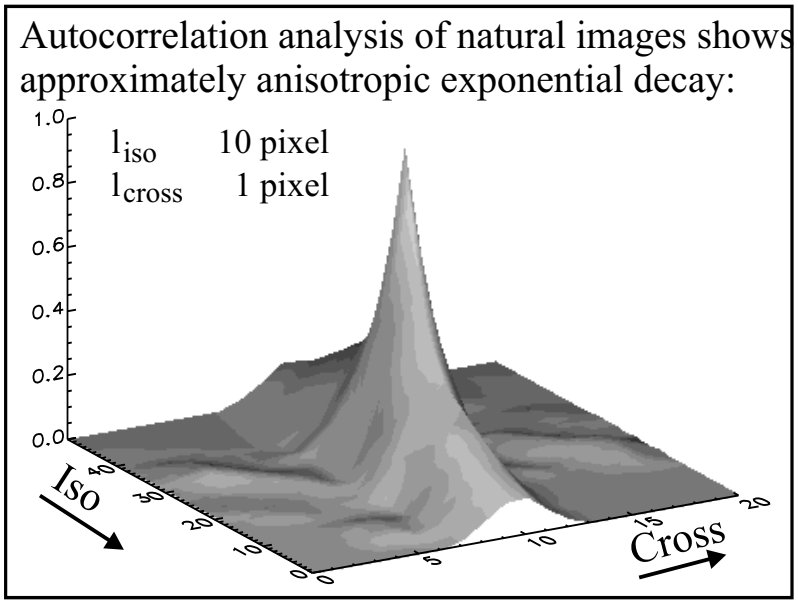


Abbildung 6.6: Räumliche Korrelation zwischen den Kantenstärken im Beispielbild. In Verlängerung der Vorzugsrichtung einer lokal detektierten Kante treten gleichartige weitere Kanten viel häufiger auf als quer dazu.

Im Beispiel ergibt sich ein Unterschied in der typischen Korrelationslänge von etwa einer Größenordnung, also ein Verhältnis der Abfallkonstanten von $\kappa = l_{iso}/l_{cross} \approx 10$. Diese Anisotropie kann als Grundlage für die Gestaltung der lateralen Kopplung genutzt werden: Auch hier wurde eine anisotrop exponentiell mit dem Abstand abfallende Kopplungsstärke gewählt.

$$w_{ij} = w_0 \cdot e^{-\sqrt{(\frac{\Delta_{iso}}{k_{iso}})^2 + (\frac{\Delta_{cross}}{k_{cross}})^2}} \tag{6.18}$$

mit

$$k_{iso} = k_0 \qquad \text{und} \qquad k_{cross} = \frac{k_0}{\kappa} \tag{6.19}$$

wobei k_{iso} und k_{cross} die beiden Längenkonstanten des Abfalls der Verbindungsgewichte und Δ_{iso} und Δ_{cross} den jeweiligen Abstand von Neuron i und j in Iso- bzw. Cross-Orientierung der Kante bezeichnen. w_0 parametrisiert die relative Gewichtung des Kopplungsinputs gegenüber dem originären RF-Input und entspricht damit dem Verhältnis von Linking- zu Feeding-Verstärkung beim Marburger Modellneuron (s. auch Abschnitt 3.3.5). k_0 kontrolliert bei konstantem Anisotropiefaktor κ die Reichweite der lateralen Kopplung.

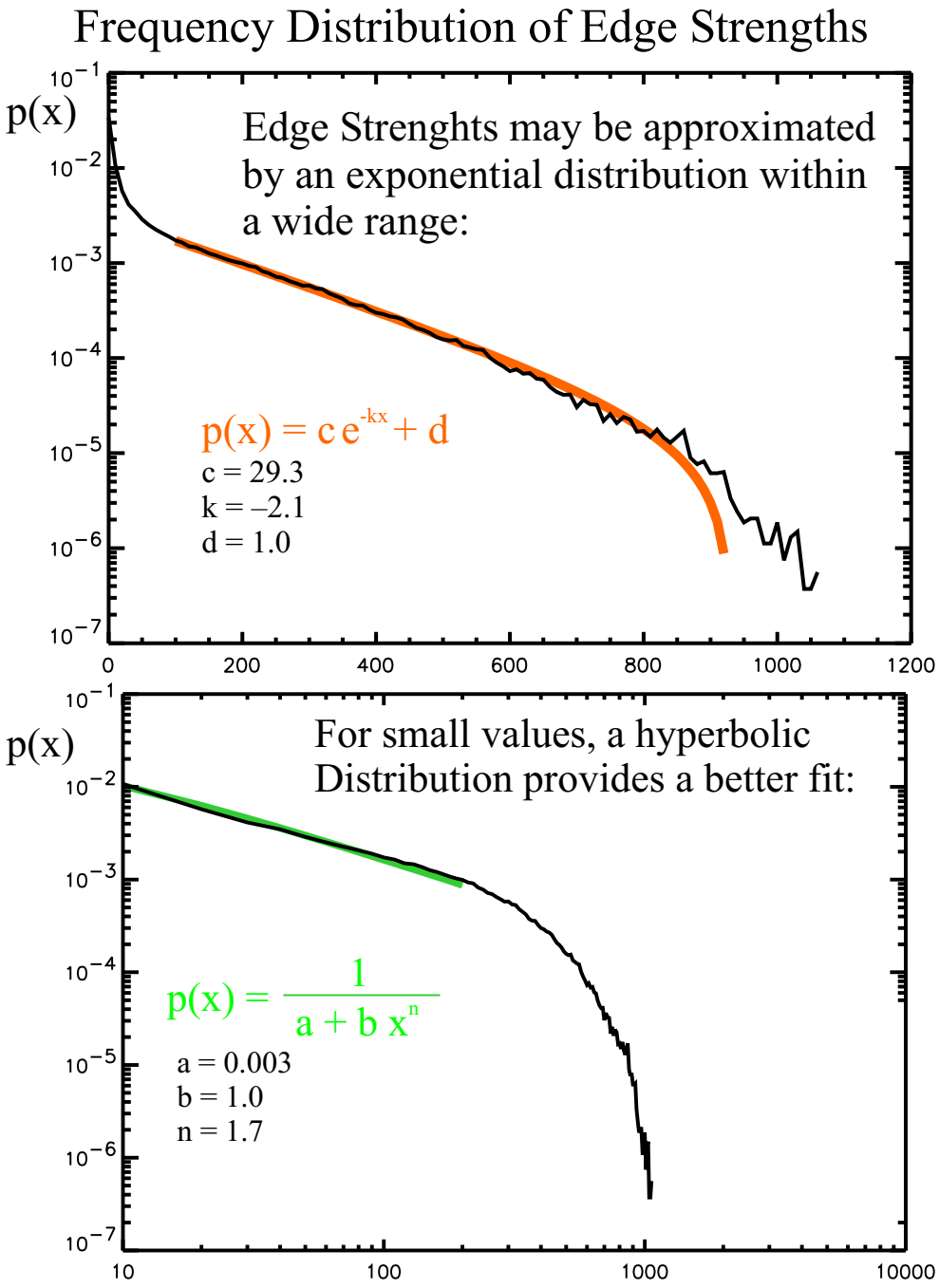


Abbildung 6.7: Häufigkeitsverteilung der Kantenstärken, die durch Faltung mit den orientierten RFs der Kantendetektoren entstanden sind. Die analytische Näherung wurde jeweils im eingezeichneten Bereich numerisch an die Kurve angefittet.

6.6.2 Ergebnisse

Abb. 6.8–6.11 (nächste Seite) zeigen die Ergebnisse für die Anzahl N_{Err} falscher Detektorantworten für beide Störungsarten und Kopplungsvarianten. Im Fall des Gaußschen weißen Rauschens ergibt sich keine Verbesserung durch eine exzitatorische Nachbarschaftskopplung. Dies war auch nicht zu erwarten, da die Störung hier ja symmetrisch wirkt. Im Fall der 'nebelartigen' Störung ist dies anders; hier wird im besten Fall eine Reduzierung der Fehlantworten um bis zu 10 % erreicht. Wie oben beschrieben bilden hier zwei gegenläufige Tendenzen ein Optimum: Je stärker die Kopplung, desto höher ist die Zahl der von der Störung unterdrückten Neurone, die durch die Kopplung wieder überschwellig werden. Gleichzeitig werden aber auch Neurone angeregt, die im ungestörten Fall unterschwellig bleiben, wodurch die Zahl der Fehler wieder ansteigt.

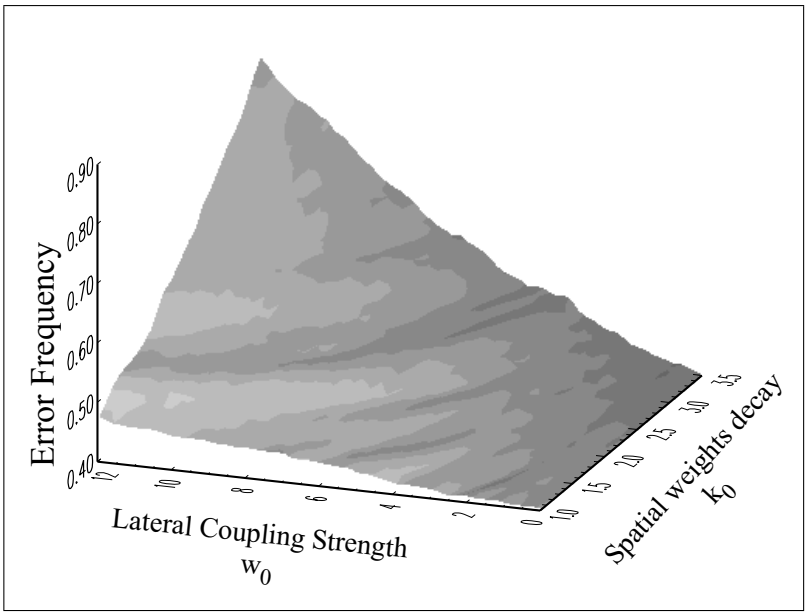


Abbildung 6.8: Auswirkung der additiven lateralen Kopplung auf den Anteil der Fehlantworten beim normalverteilten Rauschen (Abb. 6.5b). Aufgrund der symmetrischen Natur der Störung werden die Fehlantworten durch die exzitatorische additive Kopplung nicht reduziert.

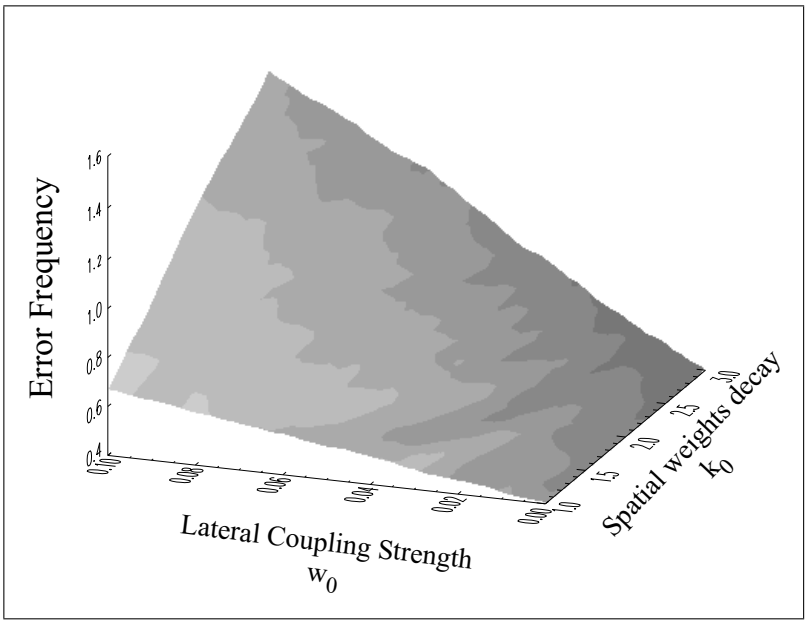


Abbildung 6.9: Auswirkung der multiplikativen lateralen Kopplung auf den Anteil der Fehlantworten beim normalverteilten Rauschen (Abb. 6.5b). Auch hier wird die Anzahl der Fehlantworten nicht verringert.

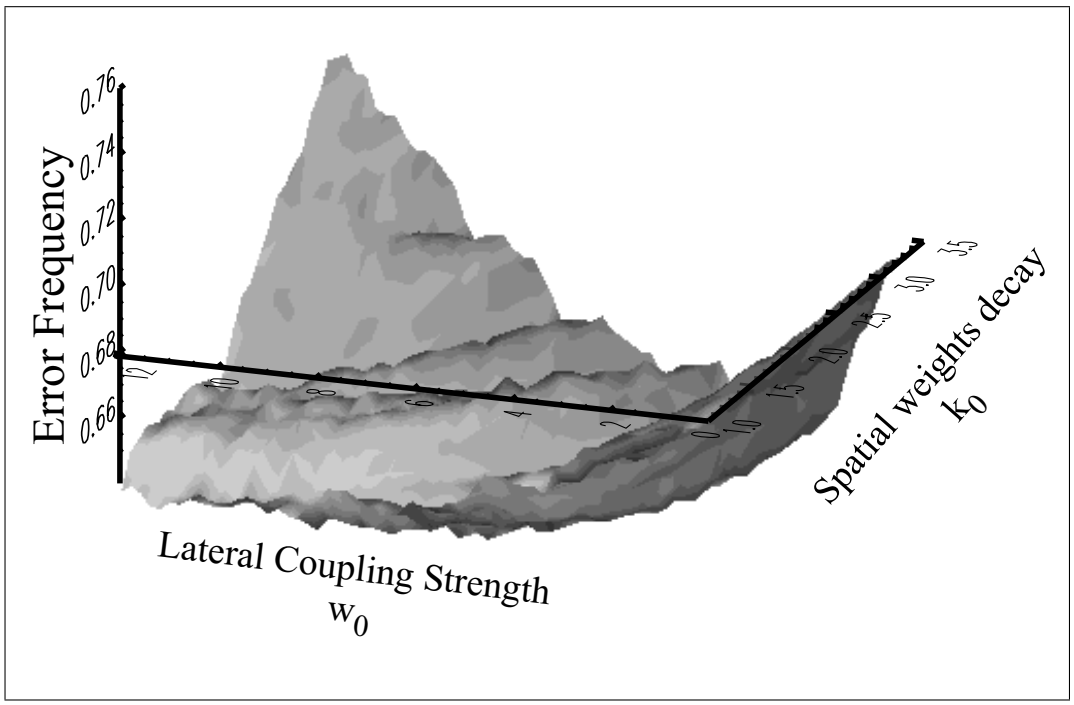


Abbildung 6.10: Auswirkung der additiven lateralen Kopplung auf den Anteil der Fehlantworten beim 'Nebel'-Rauschen (Abb. 6.5c). Die Anzahl der Fehlantworten geht bei geeigneter Parametereinstellung um bis zu 10% zurück.

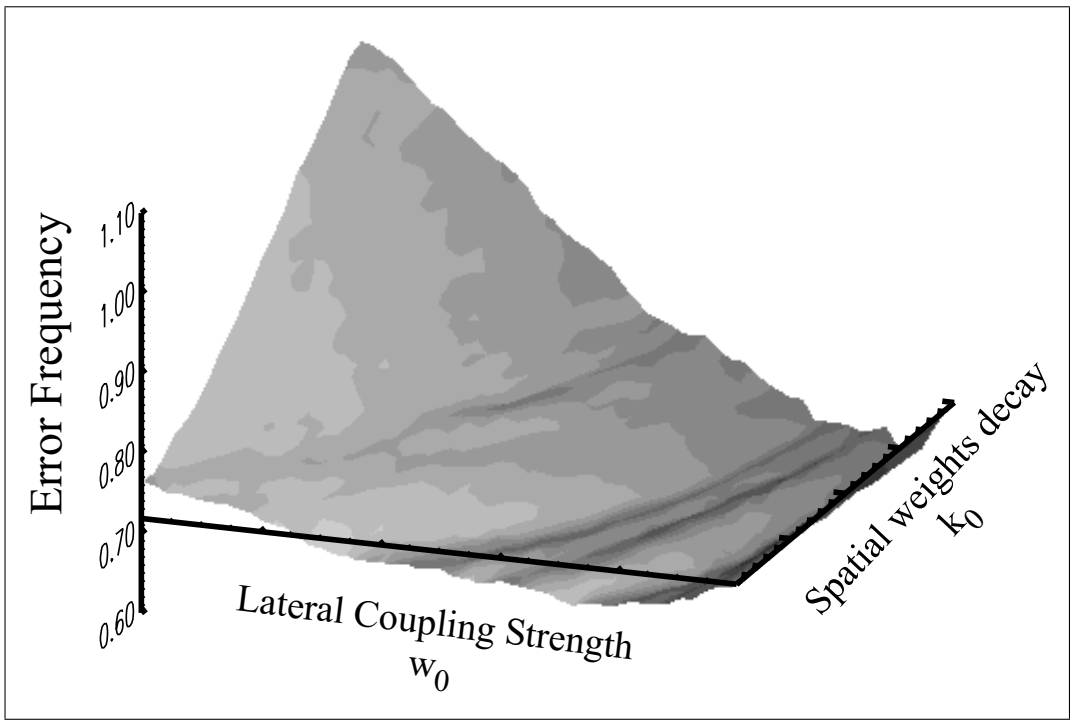


Abbildung 6.11: Auswirkung der multiplikativen lateralen Kopplung auf den Anteil der Fehlantworten beim 'Nebel'-Rauschen (Abb. 6.5c). Wie von der Theorie vorhergesagt, ergibt sich eine vergleichbare Fehlerreduktion wie bei additiver Kopplung.

7 Zusammenfassung und Diskussion

7.1 Zusammenfassung

In der vorliegenden Arbeit wurde ein neuronales Netz vorgestellt, das eine einfache Aufmerksamkeits- und Blicksteuerung in Anlehnung an physiologische Befunde verwirklicht. Der Begriff 'einfach' bezieht sich dabei darauf, daß das System rein datengetrieben arbeitet. Außer wenigen, sehr allgemeinen Grundannahmen ist weder für das Auffinden noch für die Verfolgung von Objekten spezifisches Wissen notwendig.

Im Vordergrund stand bei der Entwicklung und Zielsetzung die Funktionsfähigkeit des Gesamtsystems unter möglichst allgemeinen realen Bedingungen, d.h. bei der Anwendung auf bewegte reale Szenen. Wie die in Kap. 5 vorgestellten Simulationsbeispiele zeigen, darf diese Grundanforderung als erfüllt angesehen werden. Im folgenden versuche ich eine genauere Einordnung der Eigenschaften und Ergebnisse sowohl im Vergleich zu bisherigen technischen Systemen als auch hinsichtlich einer eventuellen physiologischen Relevanz.

7.2 Vergleich mit modellbasierten technischen Systemen zur Objektverfolgung

Im Vergleich mit den klassischen Systemen zur technischen Bildverarbeitung (s. auch Kap. 1) werden die spezifischen Vor- und Nachteile des hier verwendeten Verfahrens deutlich. Charakteristisch für die klassischen Verfahren sind folgende Eigenschaften:

- Bei der Lokalisation und Verfolgung bekannter Objekte arbeiten klassische, an die jeweilige Aufgabenstellung angepaßte Verfahren sehr genau, teilweise im Sub-Pixel-Bereich.
- Wird bei der Modellentwicklung entsprechender Aufwand getrieben, so lassen sich auch komplexe reale Szenen mit diesen Verfahren bearbeiten.
- Generell muß aber für jeden Objekttyp eine entsprechende Modellklasse entwickelt werden, d.h. neue Objekte müssen explizit implementiert werden.

Demgegenüber haben natürliche Systeme wenig Probleme, mit wechselnden Umgebungsbedingungen und unbekannten oder noch nicht klassifizierten Objekten zurechtzukommen. Im Gegenteil ist diese Situation in einer natürlichen Umgebung eher die Regel.

Eine solche Allgemeinheit in der Analyse und entsprechend flexibles Verhalten wäre auch für technische Systeme wünschenswert.[1] Das hier vorgestellte neuronale Netz wurde genau unter dieser Prämisse der Allgemeinheit entwickelt; es zeichnet sich durch folgende Eigenschaften aus:

- Es ist nur minimales Vorwissen über die aufzufindenden Objekte erforderlich: Sie müssen räumlich zusammenhängen und sich durch ihre Bewegung vom Hintergrund abheben (wobei die Merkmalsdimension 'lokale Bewegung' ggf. durch eine andere Vorverarbeitung ersetzt oder ergänzt werden kann).
- Die Auswahldynamik des Systems verhält sich robust: Es wird immer ein eindeutiges Blickziel ausgewählt und mit einer (einstellbaren) Trägheit beibehalten.
- Alle aufgefundenen Objekte können auch verfolgt werden; diese Verfolgung läßt sich in einer linearen Näherung analytisch behandeln (vgl. Kap. 3.5.8 und [Pauly et al., 1999]).
- Die Allgemeinheit beim Auffinden und Verfolgen von Objekten wird zum Teil auf Kosten der Genauigkeit erreicht; eine hochpräzise Lageschätzung ist ohne spezielles Objektwissen nicht möglich.

Ein Ansatz, das Auffinden und Erkennen von Objekten in Bildern auch bei der Anwendung von Matching-Verfahren flexibler zu gestalten, ist die Formulierung allgemeinerer Anpassungsvorschriften. Beispielsweise stellten Drüe et al. [1994] im Rahmen des NAVIS-Projekts einen Fovealisierungsalgorithmus vor, der auf Symmetrieeigenschaften von Objekten basiert. Damit lassen sich technische Objekte wie Verkehrsschilder unabhängig vom Hintergrund gut auffinden, da diese fast immer ausgeprägte Symmetrieeigenschaften aufweisen. Ein solcher Ansatz ist deutlich weniger spezifisch als ein explizit modellbasiertes Verfahren, deckt aber trotzdem nur einen relativ begrenzten Bereich von Problemstellungen ab.

Zusammenfassend läßt sich sagen, daß sich die Fähigkeiten von modellbasierten und datengetriebenen Systemen komplementär verhalten: Die datengetriebenen Systeme nach biologischem Vorbild eignen sich gut zum erstmaligen Auffinden und Verfolgen von Objekten, und zwar unabhängig davon, ob bereits Wissen über das jeweilige Objekt vorhanden ist. Auch eine einfache Segmentierung ist ohne spezifisches Vorwissen möglich. Eine genaue Analyse im Sinn einer exakten Identifizierung sowie Orts-, Lage- und Bewegungsschätzung erfordert zur Zeit dagegen weiterhin den Einsatz modellbasierter Algorithmen; nur so läßt sich die für industrielle Zwecke notwendige Genauigkeit erreichen.

7.2.1 Ausblick: Hybrid-Systeme

Betrachtet man die spezifischen Vor- und Nachteile der verschiedenen Ansätze, liegt es nahe, bei der Weiterentwicklung solcher Systeme beide Ansätze zu kombinieren, um sowohl

[1] Selbstverständlich profitieren auch Lebewesen von detailliertem Objektwissen: Wenn wir einen Radfahrer sehen (und als solchen klassifizieren können), haben wir eine gute Vorstellung von den Bewegungsmöglichkeiten und zu erwartenden Reaktionen dieses Verkehrsteilnehmers. Insofern verlassen sich technische und natürliche Systeme hier anscheinend auf eine vergleichbare Funktionalität.

ein robustes Verhalten unter schwierigen bzw. unbekannten Bedingungen als auch eine genauere Analyse bei anwendbarem Objektwissen zu erreichen.

Ein solches Hybrid-System müßte vermutlich folgendermaßen arbeiten: Eine datengetriebene Vorstufe stellt eine Vorauswahl (Aufmerksamkeitsbereich) der Eingangssequenz bereit, die unter möglichst verschiedenen Bedingungen brauchbare, robuste Ergebnisse liefert. Diese dürfen dabei durchaus ungenau sein – es muß lediglich gewährleistet sein, daß nachfolgende Stufen einen Input von deutlich reduzierter Komplexität erhalten. Unter diesen Bedingungen können klassische *Matching*-Verfahren dann wieder sehr effizient arbeiten, da der Suchraum gegenüber einem Vollbild erheblich eingeschränkt ist.

Eine besondere Rolle spielt in diesem Zusammenhang die Invarianz der Erkennung gegenüber Orts- und Lage- und Größenänderungen. Ein Hauptgrund für den riesigen Suchraum, der bei realen Szenen entsteht, ist ja gerade die Vielfalt der möglichen Erscheinungsbilder bei einem einzelnen Objekttyp. Zumindest die Lageinvarianz läßt sich durch eine geeignete Blicksteuerung wie die hier vorgestellte weitgehend sicherstellen; die verbleibende Unsicherheit bei der Ortsbestimmung bzw. Fixation ist vernachlässigbar gegenüber dem Absuchen des ganzen Sichtfeldes. Rotations- und Größeninvarianz lassen sich z.B. durch log-polare Abbildungen erzeugen [Reitböck und Altmann, 1984]. Ob derartige Transformationen für eine konkrete Aufgabestellung nützlich sind, muß für jede Anwendung getrennt entschieden werden. Im technischen Bereich sind zumindest Anwendungen denkbar, bei denen mit Hilfe solcher invarianter Abbildungen schnell entschieden werden kann, ob ein gesuchtes Objekt überhaupt im Bild vorhanden ist.

Unter dem Gesichtspunkt, allgemeine Funktionalität nach biologischem Vorbild zu erreichen, stellt allerdings die nachgeschaltete Ankopplung eines Assoziativspeichers noch einen interessanteren Weg dar. Dieser implementiert zwar auch eine Ähnlichkeitsprüfung zwischen gespeichertem Muster und präsentierter Szene, kann aber im Gegensatz zu einem explizit programmierten *Matching*-Algorithmus neue Objekte selbständig lernen. Gerade bei der Analyse bewegter Bilder wäre dies interessant; charakteristische Bewegungsmuster haben für Menschen einen hohen Wiedererkennungswert.

Tatsächlich existieren bereits funktionsfähige Systeme mit hybridem Design: Das bereits erwähnte NAVIS verwendet nach der Fovealisierung aufgrund von Symmetrieeigenschaften ein lernfähiges Matching-Verfahren, also einen Assoziativspeicher, zur Erkennung von Objekten. Auch hier trägt die Lageinvarianz durch Fovealisierung der Objekte maßgeblich zur Funktionsfähigkeit des Gesamtsystems bei.

Das von Lindemann et al. [1998] vorgestellte neuronale Netz realisiert eine modellbasierte Objektverfolgung mit der gleichen retinotop organisierten Aufmerksamkeitsschicht wie sie auch im vorliegenden Modell verwendet wird, allerdings ohne spikende Neurone. Diese wird von einem Korrelationsnetzwerk unterstützt, das ständig die Ähnlichkeit zwischen gesuchtem Objekt und dem fovealen Bereich des Gesichtsfelds prüft. Ist ein bereits gelerntes Objekt erst einmal 'eingefangen', d.h. hinreichend genau fovealisiert, läßt sich eine deutlich präzisere Verfolgung realisieren als in der datengetriebenen Variante allein.

Einen umfassenden Ansatz, biologisch motivierte und technische Verfahren in einem Gesamtsystem zu integrieren, stellen Hartmann et al. [1999] vor. Die Autoren diskutieren ausführlich die Vor- und Nachteile der verschiedenen Herangehensweisen und zeigen

Wege zur effizienten Nutzung beider Ansätze auf. Insofern spiegelt diese Arbeit recht gut den aktuellen Stand der Forschung wieder.

7.3 Physiologie

7.3.1 Sakkaden

Das vorgestellte Modell stimmt nur in Teilbereichen mit dem Wissensstand über die Steuerung von Augenbewegungen im menschlichen Sehsystem überein – im natürlichen System sind die Gegebenheiten mit Sicherheit wesentlich komplexer. Abb. 7.1 zeigt einen Überblick über die vermutlich an der Auslösung und Steuerung von Sakkaden beteiligten neuronalen Strukturen, Abb. 7.3.2 ergibt ein ähnlich kompliziertes Bild für Folgebewegungen.

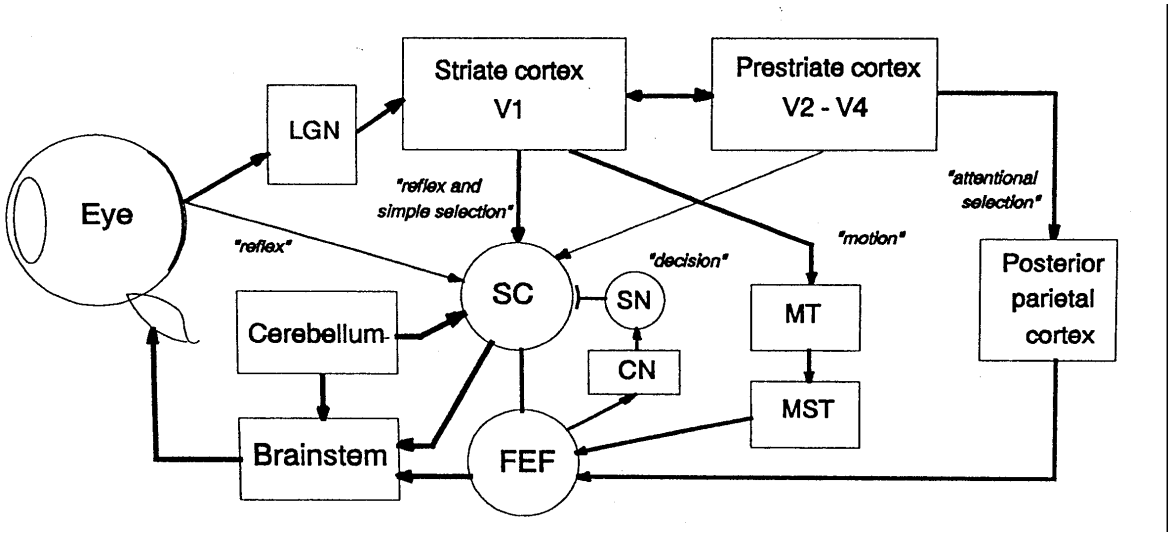


Abbildung 7.1: Vermutlich an der Steuerung von Sakkaden beteiligte Teilsysteme. (Aus: [Deubel, 1994])

Das komplexe Zusammenwirken vieler Hirngebiete ist offensichtlich erforderlich, um die vielfältigen Anforderungen an Augenbewegungen im Organismus zu erfüllen. Für die Implementation innerhalb eines – zunächst rein datengetriebenen – technischen Systems wurde der Versuch unternommen, sich auf die funktional notwendigen Teilbereiche zu beschränken. Gegenüber natürlichen Sehsystemen ist das vorgestellte Sakkadenmodell in folgenden Punkten wesentlich **vereinfacht**:

- Der gesamte Bewegungsapparat der Augen, einschließlich der *Burst-Neurone*, wurde nicht explizit modelliert; ein mechanischer Apparat wäre beim Einsatz einer realen Kamera mit zu berücksichtigen.
- Es wurden keine Top-Down-Wechselwirkungen modelliert.
- Der Input zum SC stammt lediglich aus dem Transientensystem (mit Ausnahme von Beispielszene XY); im natürlichen System können fast alle Merkmalskontraste

ein Blickziel markieren.

- Der Superior Colliculus und die frontalen Augenfelder wurden nicht getrennt modelliert; das sakkadische Netzwerk steuert zugleich die Folgebewegungen.

- Im natürlichen System sind neuronale Aufmerksamkeitskarte und Bewegungskarte vermutlich verschiedene Systeme, wodurch prinzipiell eine längere räumliche Verlagerung der Aufmerksamkeit ohne Augenbewegungen möglich wird. Dieses Phänomen wurde z.B. von Hikosaka et al. [1993a,b] elegant demonstriert; der umgekehrte Effekt (Augenbewegungen ohne Verlagerung der Aufmerksamkeit) scheint nicht möglich zu sein, vgl. z.B. [Findlay, 1999]. Interessanterweise fanden Mackeben und Nakayama [1993] für Aufmerksamkeitsverlagerungen einen *Gap Effect*, der dem sakkadischen sehr ähnlich ist und interpretierten dies als einen weiteren Hinweis, daß gezielten Augenbewegungen immer eine entsprechende Verschiebung des Aufmerksamkeitsfokus vorausgeht.

- Die Inhibition des sakkadischen Apparates (z.B. aufgrund willentlicher Fixation) wird nicht explizit modelliert; im natürlichen System wird diese vermutlich über den in Abb. 7.1 dargestellten Pfad FEF → CN → SN vermittelt (vgl. auch Munoz und Wurtz [1992]).

- Reale Sakkaden werden offensichtlich unter Einbeziehung der Geschwindigkeit des jeweiligen Blickziels berechnet; diese Fähigkeit wird durch Läsionen in Area MT stark beeinträchtigt [Newsome et al., 1985]. Groh et al. [1997] konnten bei intakter Area MT durch Mikrostimulation in dieser Region allen Typen von Augenbewegungen einen systematischen Fehler aufprägen, der sich mit einer falsch geschätzten Objektgeschwindigkeit erklären ließ.

- Die Planung von Mehrfachsakkaden ist nicht möglich; im Modell ist jede Sakkade ein in sich abgeschlossener Vorgang (vgl. z.B. [Becker und Jürgens, 1979]). Im natürlichen System wird diese Aufgabe vermutlich von den supplementären Augenfeldern (SEF) koordiniert [Pierrot-Deseilligny et al., 1995].

- Es existiert kein Verschiebungsmechanismus, der über eine Sakkade hinweg retinozentrisch codierte visuelle Reize in das neue Koordinatensystem transformiert. Derartige Zellen wurden von Mays und Sparks [1980] im SC entdeckt. Für die präzise Ausführung und Planung von Mehrfachsakkaden ist entweder eine Transformation über einzelne Sakkaden oder eine Codierung in einem kraniozentrischen Koordinatensystem erforderlich. Henriques et al. [1998] konnten unter Verwendung von *Open-Loop*-Reizen zeigen, daß ersteres bei Affen der Fall ist. Dominey und Arbib [1992] verwendeten dieses Konzept, um in einem physiologisch orientierten Modell die Planung solcher Bewegungssequenzen zu ermöglichen.

Die Rechtfertigung für diese erheblich vereinfachte Modellierung liegt einerseits in der technischen Forderung nach einfacher Umsetzung und Echtzeitfähigkeit begründet. Andererseits sind viele biologische Einzelheiten für die angestrebte Funktion aber auch gar nicht notwendig: Angestrebt wurde ein rein datengetriebenes, dafür aber robustes

Modell. Damit ist die Berücksichtigung von Top-Down-Einflüssen zunächst zweitrangig, auch wenn ihre spätere Einbeziehung beim Modellentwurf bereits berücksichtigt wurde.

Im natürlichen System fällt den FEF vermutlich in hohem Maß die Rolle eines 'Vermittlers' zwischen Bottom-Up und Top-Down-Pfad zu. Einerseits weisen sie in ihrer Funktion bei der Steuerung von Sakkaden anscheinend eine hohe Redundanz zum SC auf: Klinische und tierexperimentelle Läsionsstudien zeigten, daß durch Schädigung eines der beiden Areale die Fähigkeit zu visuell geleiteten Sakkaden noch nicht verlorengeht. Erst wenn beide Areale lädiert sind, können Sakkaden nicht mehr präzise auf ein sichtbares Blickziel gelenkt werden [Keating und Gooley, 1988; Schiller und Chou, 1998; Pierrot-Deseilligny et al., 1991, 1995].

Andererseits scheinen die FEF unabdingbar für die Auslösung nicht visuell gestützter Sakkaden zu sein: Die Fähigkeit, gedächtnisgeleitete oder Anti-Sakkaden-Aufgaben korrekt auszuführen, läßt nach einer Schädigung der FEF stark nach [Guitton et al., 1985; Burman und Bruce, 1997; Hanes et al., 1997]. Zudem sind sie offensichtlich an der Steuerung von Folgebewegungen beteiligt; dabei spielen, ähnlich wie bei nicht-visuellen Sakkaden, kognitive Prozesse generell eine wichtige Rolle.

Dies legt die Vorstellung nahe, daß die funktionalen Aufgaben geteilt werden: Der SC kann einfache, reflexive Sakkaden zu sichtbaren Zielen schnell und ohne Umweg durchführen; über die FEF können kortikale Anforderungen einfließen. Da derartige Top-Down-Einflüsse im Modell noch nicht berücksichtigt sind, entspricht es hinsichtlich der Sakkadensteuerung eher einem 'isolierten SC'. Andererseits zeigen einige Grundfunktionen bei der Auswahl von Sakkadenzielen sowie beim Wechselspiel zwischen Fixation und Sakkade bemerkenswerte **Ähnlichkeiten**:

- Munoz und Wurtz [1992, 1993a,b, 1995a,b] wiesen nach, daß das Sakkadenziel im SC wie im Modell in der Aktivität einer retinotopen Karte codiert ist. Die Rolle eines Fixationszentrums (entsprechend dem zentralen Bereich der Aufmerksamkeitskarte im Modell) hat dabei der rostrale Pol des SC; mit wachsender Entfernung vom rostralen Ende des SC wird auch der codierte Bewegungsvektor größer (entsprechend den peripheren Bereichen im Modell).

- Weitere Unterstützung für die Idee einer gegenseitigen Inhibition von Fixations- und Bewegungszellen liefert die Beobachtung von Munoz und Wurtz [1993a,b], daß es für die Auslösung einer Sakkade zwingend notwendig ist, daß die *Omnipause*-Zellen im Hirnstamm ihre Aktivität zunächst einstellen.

- Die *Bottom-up*-Gewinnung eines Aufmerksamkeitssignals aus den Grauwert-Eingangsbildern durch relativ einfache, aber nichtlineare Operationen auf der Helligkeitsverteilung wird auch aus Experimenten berichtet. Findlay et al. [1993] fanden, daß bei der Verwendung von schachbrettartigen Reizen der *Betrag* des Intensitätskontrastes und die visuelle Ausdehnung den 'Aufmerksamkeitswert' eines Blickziels bestimmt. Dieser Effekt wird im Modell durch die gleichberechtigten Summation von ON- und OFF-Kanälen in der Vorverarbeitung erzielt.

- Dias und Bruce [1994] wiesen nach, daß ein Teil der FEF-Neurone, von denen bekannt ist, daß sie eigentlich Sakkadenziele repräsentieren, regelmäßig bereits vor

Beginn einer Sakkade aktiv wird. Erstaunlicherweise feuern sie sogar, wenn die (angekündigte) Sakkade gar nicht ausgeführt wird. Sie folgern daraus, daß diese Neuronen eine aktive Rolle beim Aufheben einer Fixation spielen – was genau der Inhibition zwischen peripheren und zentralen Bereichen der Aufmerksamkeitsschicht im Modell entspricht.

- Gottlieb et al. [1994] fanden in der *Area lateralis interparietalis* (LIP) von Affen Neurone, die nur dann auf Reize innerhalb ihres RF ansprachen, wenn diese entweder transient eingeblendet wurden oder für die Lösung der jeweiligen Aufgabe relevant waren. Genau dieses neuronale Verhalten wird aber für die Konstruktion einer Aufmerksamkeitskarte gebraucht. (Die Area LIP beim Affen entspricht von der Funktion her den parietalen Augenfeldern beim Menschen, vgl. [Pierrot-Deseilligny et al., 1995]). Im vorgestellen Modell wird diese Karte zwar lediglich aus einfachen sensorischen Merkmalen gewonnen (Intensitätskonstrast bzw. Bewegungskontrast); die Top-Down-Einflüsse fehlen. In beiden Fällen findet jedoch eine starke Vorselektion des Inputs statt, auf deren Grundlage sich das wichtigste Sakkadenziel ermitteln läßt.

Ein auch in der Literatur nicht völlig geklärter Punkt ist die Frage, wie die *räumliche* Berechnung des Sakkadenziels und die *zeitliche* Auslösung ineinandergreifen. Den ersten direkten Hinweis, daß beide Parameter eventuell getrennt erzeugt werden, gaben die Doppelsakkaden-Experimente von Becker und Jürgens [1979]. Der dabei gefundene, annähernd lineare Zusammenhang zwischen der (ab der zweiten Sakkade berechneten) Reaktionszeit und Zielort bei schnellen, visuell geführten Doppelsakkaden wurde von den Autoren als *Amplitude Transfer Function* bezeichnet. Die Autoren führten ihn darauf zurück, daß die von der Auslösezeit abhängige Sprungweite die noch laufende Planung des zweiten Teils der Doppelsakkade widerspiegelt (das 'Durchfahren' der Funktion kam ohne experimentelles Zutun allein durch die stochastische Variabilität der Reaktionszeiten zustande).

Folgt man dieser These, so scheint eine Trennung von Zielplanung und Reaktionsauslösung zunächst unumgänglich. Allerdings können Modelle wie das hier vorgestellte die beiden Prozesse doch wieder auf einen einzigen zurückführen: Die Sakkade wird ausgelöst, wenn der Schwerpunkt der Aktivität der Aufmerksamkeitsschicht einen genügend großen Fixationsfehler anzeigt, was auch der Fall sein kann, während die Aktivität noch vom alten Blickziel auf ein neues übergeht (vgl. z.B. den Übergang zwischen den Blickzielen in Abb. 5.7, Zeitschritte 199 und 225). Natürliche sakkadische Reaktionszeiten zeigen auch bei identischen Reizbedingungen eine große statistische Variabilität [Kowler, 1990]. Enthält die Signalkette ein stochastisches Verzögerungsglied, so kann der Auslösezeitpunkt praktisch jedes Zwischenstadium einer laufenden Kompetitionsdynamik treffen. Eine ausführliche Diskussion dieser Problematik mit ähnlichen Schlußfolgerungen findet sich bei Findlay [1999], zusammen mit einem ebenfalls funktional orientierten Modellvorschlag. Ergänzend dazu fanden Findlay und Gilchrist [1997], daß Einflüsse aus höheren Ebenen (in diesem Fall eine Unterscheidung Kreis/Quadrat) sich nur bei den langen Reaktionszeiten einer experimentellen Verteilung auf das Sakkadenziel auswirken. Auch dies ist ein Indiz, daß eine hinreichende Aktivität in der Aufmerksamkeitskarte eine

Sakkade 'automatisch' auslösen kann, die dann rein visuell gestützt und unempfindlich gegenüber Einzelheiten des Stimulus ist.

Generell scheint auch ein stark vereinfachtes Modell der Forderung nach einem robusten und sinnvollen Verhalten bei der *datengetriebenen* Auswahl von Blickzielen zu genügen. Einige wichtige Grundfunktionen natürlicher Sehsysteme können dabei nachgebildet und praktisch eingesetzt werden, ohne daß man die gesamte Komplexität der biologischen Systeme benötigt.

7.3.2 Folgebewegungen

Was Folgebewegungen angeht, ist die Situation im biologischen System vermutlich noch komplexer als bei der Sakkadensteuerung, was sich auch in teilweise widersprüchlichen Standpunkten in der Literatur niederschlägt. Weitgehende Einigkeit herrscht darüber, daß die zugehörige Sensor-Motor-Schleife in erster Linie Geschwindigkeitssignale und nicht Ortssignale verarbeitet. Dies wurde zum erstenmal von Rashbass [1961] unter Verwendung sogenannter *Step-Ramp*-Stimuli gezeigt: Ein isolierter Lichtpunkt springt vom Zentrum auf eine leicht periphere Position und beginnt sofort, sich wieder auf das Zentrum zuzubewegen. Die Reaktion auf diesen Reiz ist die gleiche, die auch ohne den anfänglichen Sprung zu beobachten ist: eine Folgebewegung, die nach einer Beschleunigungsphase annähernd die Geschwindigkeit des Blickziels erreicht, so daß dieses in einer fast konstanten, aber nicht zentralen Position auf der Retina zu sehen ist. Dies ist ein wesentlicher Unterschied zum hier vorgestellten Modell, bei dem auch die Steuerung der Folgebewegungen hauptsächlich auf der Auswertung eines Positionssignals beruht (das verfügbare Geschwindigkeitssignal wird nur ergänzend hinzugezogen, vgl. Kap. 3.5.8.4). Die folgende, aus [Gottlieb et al., 1994] entnommene Abb. 7.3.2 zeigt die vermutlich an der Steuerung von Folgebewegungen beteiligten Hirnareale.

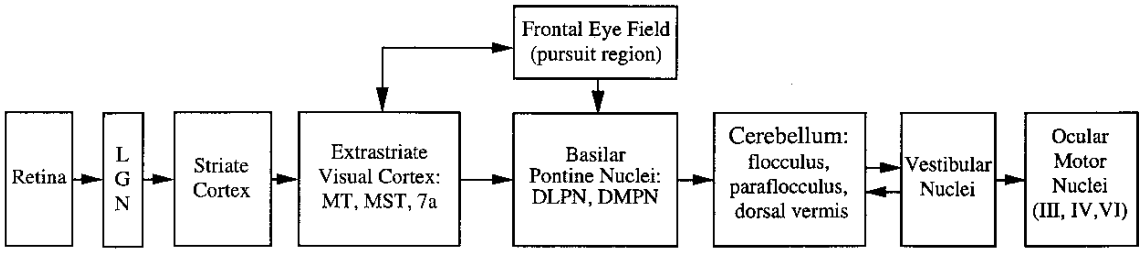


Abbildung 7.2: Übersicht über die vermutlich an der Steuerung von Folgebewegungen beteiligten Hirnareale. (Aus: [Gottlieb et al., 1994])

Bei der Untersuchung von Folgebewegungen an Hasen zeigte sich, daß diese über ein vergleichsweise einfaches Verfolgungssystem verfügen, das sich weitgehend durch einen linearen Regelkreis auf der Basis von Geschwindigkeitssignalen beschreiben läßt, vgl. z.B. [Ter Braak, 1936; Collewijn, 1972, 1981]. Einige der Ergebnisse aus dieser Forschung lassen sich auch auf den Menschen übertragen, so z.B. die Eigenschaft, daß zur Aufrechterhaltung einer Folgebewegung ein gewisser retinaler Schlupf benötigt wird [de Pucket

und Steinman, 1969; Kowler und McKee, 1987]. Unterschiede zeigen sich allerdings, wenn man versucht, mit künstlich stabilisierten Blickzielen die Arbeitsweise der 'offenen Regelschleife' zu untersuchen: Hasen verhalten sich, als wäre das Blickziel nicht mehr vorhanden [Collewijn und van der Mark, 1972], die Blickrichtung bleibt bis auf kleine Schwankungen stabil. Menschen zeigen dagegen ein komplexes Spektrum von Verhaltensweisen, das sowohl Stabilität als auch willentlich ausgelöste glatte Bewegungen beinhaltet; zudem unterscheiden sich die beobachteten Bewegungsmuster stark zwischen den Versuchspersonen [Heywood, 1972].

Kowler [1990] kommt zu dem Schluß, daß, anders als beim Hasen, eine Beschreibung der Steuerung von Folgebewegungen auf der Grundlage eines einfachen systemtheoretischen Ansatzes überhaupt nicht sinnvoll möglich ist. Ihrer Meinung nach sind kognitive Prozesse und visuelle Objektverfolgung untrennbar verbunden. Um diesen Standpunkt zu untermauern, führt sie u.a. die vielfältigen Typen von antizipatorischen Folgebewegungen bei bekannter wie unbekannter zukünftiger Targetbewegung [Kowler und Steinman, 1979a,b, 1981] sowie das Verhalten bei mehrfachen Blickzielen an [Dodge und Fox, 1928; Collewijn und Tamminga, 1986]. Situationen mit mehreren Blickzielen sind besonders unter funktionalen Gesichtspunkten von großem Interesse, da sie einer natürlichen visuellen Reizsituation sicherlich viel näher kommen als isolierte Punktreize. Das typische Verhalten in einem solchen Fall ist, daß es nur wenig oder gar keine Überlagerung zwischen verschiedenen möglichen Blickzielen gibt, sondern daß eines der Blickziele sehr genau verfolgt bzw. fixiert wird, während die anderen ignoriert werden. Dies ist sogar der Fall, wenn großflächige Streifen- oder Punktmuster sich bei verschiedener Bewegungsrichtung gegenseitig durchdringen [Ter Braak, 1957; Ter Braak und Buis, 1970; Kowler et al., 1984]. Die Entscheidung für eines von mehreren möglichen Verfolgungszielen wird allgemein als typisches Phänomen der selektiven Aufmerksamkeit gesehen. Aus funktionaler Sicht ist diese Fähigkeit entscheidend, um in einer komplexen, natürlichen Umwelt mit ihren vielfachen Bewegungsreizen einzelne bewegte Objekte zu fixieren, zu verfolgen und schließlich (soweit möglich) zu identifizieren (vgl. 3.5.1). Genau diese Aufgabe erfüllt die in dieser Arbeit vorgestellte Aufmerksamkeitssteuerung aber in robuster Weise: Aus vielen möglichen Blickzielen wird eines ausgewählt, so daß eine Verfolgung ermöglicht wird. Auch hier sind beim Menschen kortikale Prozesse mit Sicherheit beteiligt; so können Versuchspersonen z.B. auch willentlich zwischen verschiedenen Reizen 'umschalten' [Dubois und Collewijn, 1979] – eine Leistung, die einem datengetriebenen Modell nicht ohne weiteres möglich ist.

Es bleibt festzuhalten, daß das vorgestellte Modell in einem Punkt wesentlich vom biologischen Vorbild abweicht: Folgebewegungen werden aufgrund eines Positions- und nicht aufgrund eines Geschwindigkeitssignals gesteuert. Für die Funktion ergeben sich dadurch nicht unbedingt Nachteile, im Gegenteil sind einige Punkte für die technische Umsetzung sogar als vorteilhaft anzusehen:

- Das System relaxiert immer zu einer Situation, bei der das Blickziel sich auch im Blickzentrum befindet. Dies ist für ein technisches System sicherlich einfacher zu bewältigen als die Verfolgung (und ggf. Erkennung) von Objekten in peripherer Lage.

- Bei der Verwendung spezialisierter Hardware ist es von großem Vorteil, wenn keine getrennten Subsysteme für Sakkaden und Folgebewegungen benötigt werden; die Neuronenzahl stellte im Hardware-Accelerator eine der wichtigsten Einschränkungen dar (vgl. Kap. 2.6.2).
- Das Geschwindigkeitssignal wird zusätzlich ausgewertet; die darin enthaltene Information wird für eine schnelle Anpassung an beginnende bzw. beschleunigte Bewegungen des Blickziels verwendet.

7.4 Segmentierung

Das in Kap. 4 besprochene Verfahren zur Bildsegmentierung durch neuronale Dynamik stellt eine direkte Umsetzung der Synchronisationshypothese in ein relativ einfaches Modell dar. Wie die dort präsentierten Ergebnisse und – in größerem Umfang – die Arbeiten von Spengler [1996] und Weitzel et al. [1997] sowie Weitzel [1998b] zeigen, ist damit die Lösung einfacher Segmentierungsaufgaben auch bei realen, bewegten Szenen möglich. Voraussetzung ist allerdings, daß die Vorverarbeitung in gewissem Umfang auf die vorkommenden Objekte zugeschnitten ist: Ein konturbasiertes System wie das hier verwendete kann am besten Objekte mit fortlaufenden Rändern im Grauwertbild segmentieren; komplexere, über andere Merkmale definierte Zusammenhänge müßten über eine Erweiterung der Vorverarbeitung eingebunden werden. In den genannten Arbeiten werden *Konturecken* explizit mit berücksichtigt; Weitzel [1998a] verwendet auch texturbasierte Segmentierung. In allen diesen Arbeiten sind implizit auch die biophysikalisch begründeten Latenzzeiten der neuronalen Signale aufgrund unterschiedlichen Kontrastes enthalten. Die vorliegende Arbeit integriert darüber hinaus die *aktive* Erzeugung bzw. Nutzung solcher Latenzen im Zusammenhang mit visueller Aufmerksamkeit. Dieses Konzept steht einerseits im Einklang mit neurophysiologischen und psychophysischen Erkenntnissen zur fokalen Aufmerksamkeit; andererseits erleichtert es dem verwendeten Segmentierungsnetz die Aufgabe, indem es ihm den Input in zeitlich vorsegmentierter Form darbietet. Im Verein mit einer *aktiven Kamera* läßt sich zudem die Komplexität der Segmentierungsaufgabe durch gezielte Bearbeitung der interessanten Bildausschnitte deutlich reduzieren. Auch das ist eine wesentliche Erweiterung gegenüber den o.g. Arbeiten: Die Segmentierungsdynamik arbeitet um so stabiler, je weniger Objekte im Bild vorhanden sind.

Trotz dieser prinzipiellen Lösbarkeit einfacher Segmentierungsaufgaben auch bei bewegten Bildern ist das verwendete Modell aus folgenden Gründen kritisch zu beurteilen:

- Die Dynamik ist auch bei geeigneter Wahl der Parameter zu wenig robust, um mit der Vielzahl von Objekten und ihren Erscheinungsbildern in realen Szenen umgehen zu können.
- Eine mögliche Rückkopplung von Objektwissen in ein solches Netz wurde bisher nur in [Stöcker, 1993] mit rückwirkenden Linking-Synapsen gezeigt. Dieser an sich sehr interessante Ansatz dürfte für eine robuste Verarbeitung realer Szenen aber nur schwierig einzusetzen sein, da er auf dem präzisen Einhalten des Spike-Timings zwischen verschiedenen Neuronengruppen beruht.

- Eine mehrstufige Codierung komplexer Objekte ist mit der verwendeten Dynamik nicht möglich, allenfalls kann das System mehrere 'Segmentierungs-Vorschläge' durch die gleichzeitige Verwendung mehrerer Auflösungsstufen liefern. Aus funktionaler Sicht wäre aber eine flexiblere Behandlung von 'Zusammengehörigkeit' wünschenswert.

- Gegenphasige Oszillationen benachbarter Neuronengruppen sind im Gehirn bisher nicht gefunden worden. Neuere Ergebnisse aus der Neurobiologie deuten zwar darauf hin, daß Neurone, die ähnliche, räumlich eng benachbarte Reize codieren, in ihrer Aktivität bis zu einem gewissen Grad korreliert sind. Ab einem kortikalen Abstand von etwa 5 mm wurde jedoch keine Phasenkopplungen mehr beobachtet [Jürgens et al., 1999].

Vor allem die beiden letzten Punkte verdienen genauere Betrachtung: Ein datengetriebenes, *Bottom-Up* arbeitendes Segmentierungssystem erscheint zwar einerseits notwendig, um die enorme Komplexität des visuellen Datenstroms bereits am Anfang der Verarbeitung soweit wie möglich zu reduzieren. Andererseits darf ein solches (Teil-)System noch nicht über ausgeprägtes Objektwissen verfügen; dies stünde im Widerspruch zur Forderung der Allgemeinheit.

Sinnvoller als – wie im hier verwendeten Modell – eine einzige Segmentierung anzubieten, wäre aus funktionaler Sicht eine kontinuierliche Einordung der Bildorte nach 'Wahrscheinlichkeit der Zusammengehörigkeit'. Idealerweise sollten aus einer solchen Einordung Aussagen resultieren wie: Das Linienelement auf 10° links, 5° oben gehört mit 80% Wahrscheinlichkeit zum gleichen Objekt wie das Linienelement auf 6° links, 5° oben.

Wie aber könnte eine solche Aussage in neuronaler Aktivität codiert sein? Das hier verwendete Segmentierungsnetz ist sicher weit von physiologischen Gegebenheiten entfernt; wie die Überlegungen aus Kap. 4 zeigen, schränkt insbesondere die global wirkende Inhibition die 'effektiven' Freiheitsgrade der Dynamik stark ein. Dieser Effekt war im vorliegenden Projekt erwünscht, weil er eine relativ eindeutige, zuverlässige Interpretation der neuronalen Aktivität in Termen von Segmentierung erlaubt. Ein stärker an den physiologischen Gegebenheiten orientiertes Modell sollte über eine verteilte, lokal wirkende Inhibition verfügen und Laufzeiten neuronaler Signale berücksichtigen. Derartige Modelle wurden auch bereits von vielen Autoren untersucht, vgl. z.B. [Nischwitz und Glünder, 1995; Gerstner et al., 1991].

In den meisten dieser Arbeiten ist ein Problem darin zu sehen, daß sich bei stärker physiologisch orientierter Netzwerkarchitektur eine sehr komplexe und, zumindest auf den ersten Blick, unübersichtliche Dynamik entwickelt, die eine Einbindung in technische Systeme sehr schwierig erscheinen läßt. Weiterführend erscheinen besonders diejenigen Arbeiten, in denen biologienahe Modelle wenigstens teilweise analytisch behandelt werden konnten [Gerstner et al., 1991, 1996].

Gabriel und Eckhorn [1999] berichten interessanterweise von einer direkten Messung 'laufender Aktivitätswellen' in V1, die im Prinzip mit derartigen Modellen vereinbar ist. Saam et al. [1999] präsentieren ein Modell, das mit einer sehr biologienahen Struktur eben diese Messungen gut reproduziert und eng mit denen von Ritz et al. [1994] und Kistler et al. [1998] verwandt ist.

Kennzeichnend für das Modell von Saam et al. [1999] ist die eine Dynamik, in der sich die Synchronisation von Neuronengruppen am besten in Wahrscheinlichkeiten im Sinne einer Zeitmittelung beschreiben läßt. Dabei nimmt die Wahrscheinlichkeit, Neuronengruppen synchronisiert anzutreffen, stark mit ihrer relativen Entfernung ab. (Im Grenzfall starker lateraler Kopplung und kleiner Laufzeiten ergibt sich wieder globale Synchronisation.) Sollte sich herausstellen, daß diese Wahrscheinlichkeit auf zuverlässige Weise mit den äußeren Reizbedingungen (etwa dem Abstand zweier Linienstücke) verknüpft ist, könnte sich daraus eine **Reformulierung der Synchronisationshypothese** für datengetriebene Segmentierung ergeben:

> **Verschiedene Neuronengruppen innerhalb von V1 geben durch den Grad ihrer Synchronisation eine Wahrscheinlichkeit an, daß die Merkmale, die sie codieren, zum selben Objekt gehören.**

Diese Formulierung böte zwei Vorteile:

- In den frühen visuellen Verarbeitungsschichten erfolgt noch keine endgültige Segmentierung, sondern lediglich eine 'Schätzung von Zusammengehörigkeit'.
- In Verbindung mit einem geeignet definierten Synchronisationsmaß ließe sich dieser neuronale Code aus einem simulierten neuronalen Netz im Prinzip direkt auslesen und in technische Systeme übernehmen.

In Grundlagensimulationen konnte Saam [1999] zeigen, daß das Modell aus [Saam et al., 1999] eine solche Codierung prinzipiell leisten kann: Bei Reizung mit einer aus zwei Stücken bestehenden, unterbrochenen Kante war die Aktivität der Neurone innerhalb eines Kantenstücks wesentlich stärker synchronisiert als die zwischen den Kantenstücken; der Unterschied war um so größer, je weiter die beiden Teile der Kante voneinander entfernt waren. Als Synchronisationsmaß verwendete Saam die Fläche des Peaks unter einem Teil der Kreuzkorrelation der Aktivität der betrachteten Neuronengruppen, sehr ähnlich der Berechnung des Kap. 4.1 definierten Segmentierungsindex. Allerdings ergab sich aufgrund der schwachen lateralen Kopplung und der stark verteilten, lokal wirkenden Inhibition auch innerhalb eines Kantenstückes nie eine Korrelation von 1 (vollständige Synchronisation) sondern immer deutlich darunter; zudem entstanden die Werte erst durch zeitliche Mittelung.

Die Experimente von Gail et al. [1999a,b] kommen zu ähnlichen Aussagen: Innerhalb der Repräsentation einer Kante (genauer: eines Grauwert-Sinusgitters) in V1 ist die Aktivität signifikant korreliert. Werden die Kanten des Gitters aber durch Verschieben eines Teils unterbrochen, so bricht die Signalkorrelation über die Unterbrechung hinweg ein, während sie innerhalb des Objekts erhalten bleibt. Auch hier beruht die Korrelationsmessung auf einer zeitlichen Mittelung, d.h. einzelne Momentaufnahmen auf der Millisekunden-Skala lassen noch keine Aussage über die Kopplung zu.

Diese Ergebnisse sind allerdings noch als vorläufige Hinweise anzusehen, die weiterer Prüfung bedürfen. Aus der Sicht der angewandten Bildverarbeitung wäre es aber eine interessante Aufgabe, die Brauchbarkeit und Robustheit derartiger (im stochastischen Sinn) kontinuierlicher Repräsentationen von Zusammengehörigkeit zu prüfen; dazu käme

eine Prüfung, wie gut eine Einbindung in technische Systeme gelingt. Die Synchronisationshypothese in ihrer o.g. Form könnte dabei der Schlüssel zu dieser Einbindung sein, indem sie eine erste, datengetriebene Bewertung für die Bildsegmentierung liefert, die sich explizit in Wahrscheinlichkeitsaussagen umsetzen läßt.

7.4.1 Bedingte Wahrscheinlichkeiten – Vergleich von additiver und multiplikativer Nachbarschaftskopplung

Die statistische Analyse aus Kap. 6 ergänzt diese Sichtweise, indem sie die Rahmenbedingungen (genauer: Impulswahrscheinlichkeitsdichten) für die mikroskopische Dynamik auf einer sehr viel langsameren Zeitskala untersucht. Grundannahme ist dabei, daß sich ein großer Teil unseres Erfahrungswissens durch *bedingte Wahrscheinlichkeiten* formalisieren läßt (eine umfassende Diskussion dieser Problematik geben z.B. Rieke et al. [1999]). Für die Konturdetektion führt dies zur Formulierung aus Kap. 6. Die laterale, anisotrope Kopplung von Liniendetektoren entspricht dabei dem Gestaltgesetz vom *Guten Verlauf*. Unter der Annahme, daß äußere und neuronale Vorgänge tatsächlich auf getrennten Zeitskalen zu betrachten sind, lassen sich die additive und die multiplikative Variante der Nachbarschaftskopplung nun vergleichen:

- Beide Kopplungen wirken exzitatorisch, d.h. sie erhöhen das Membranpotential des jeweiligen Zielneurons und damit seine Wahrscheinlichkeit, zu einem bestimmten Zeitpunkt überschwellig zu werden.
- Soweit es darum geht, die An- oder Abwesenheit einer Kontur zu detektieren, sind beide Kopplungstypen in gleichem Maß von Rauschen und Störungen betroffen.
- Das von Eckhorn et al. [1990] angeführte Argument, die multiplikative Kopplung wirke 'intelligenter', da nur dort eine Verstärkung eintritt, wo bereits Feeding-Input vorhanden ist, läßt sich nur noch eingeschränkt aufrechterhalten: Im verrauschten Fall kann auch eine 'verrauschte Null' irrtümlich verstärkt werden.
- Trotzdem hat die multiplikative Kopplung gegenüber der additiven Vorteile bei der Lokalisation von Linienenden: Die Wahrscheinlichkeit einer falsch überschwelligen Antwort unmittelbar neben dem Linienende ist auch im verrauschten Fall geringer als beim additiven Typ.

Die Frage, welcher Typ von lateraler Kopplung einer konkreten Anwendung zum Einsatz kommen sollte, muß also von Fall zu Fall entschieden werden. Den Vorteilen der multiplikativen Kopplung steht ein höherer Rechenaufwand bei der Implementierung gegenüber; dies kann insbesondere bei der Hardware-Implementation Schwierigkeiten bereiten. Generell sind die Auswirkungen der zusätzlichen Nichtlinearität bei der multiplikativen Kopplung nicht so gravierend wie es auf den ersten Blick erscheinen mag: Das Ausgangssignal des Neurons entsteht ja ohnehin erst aufgrund einer viel stärker nichtlinearen Operation, dem Schwellenvergleich (vgl. dazu die beiden Kennflächen in Abb. 6.4).

Im Fall der betrachteten 'Nebel'-Störung ist vermutlich die laterale Kopplung (gleich welchen Typs) bei weitem noch nicht die bestmögliche Lösung: Wie eine genauere Analyse

zeigt, werden bei stärker werdender Kopplung zuerst die Neuronen falsch überschwellig, die unmittelbar an existierende Konturen angrenzen. Dies kann man sich leicht veranschaulichen, wenn man sich die räumliche Verteilung der Membranpotentiale der Konturdetektoren als 'Gebirge' vorstellt, das durch die Verstärkung – bei gleichbleibender Schwelle – im ganzen angehoben wird. Möglicherweise ließe sich dieser Effekt durch eine Umverteilung von Signalenergie von tieferen zu höheren Ortsfrequenzen (also einer Konturschärfung) teilweise kompensieren. Das würde jedenfalls einen Teil der mit der Störung einhergehenden Tiefpaß-Filterung wieder aufheben.

7.4.1.1 Einfluß der lateralen Kopplung auf die neuronale Dynamik

Der genaue Typ der lateralen Kopplung beeinflußt jedoch nicht nur die Konturdetektion bzw. die Wahrscheinlichkeit, daß bestimmte Merkmale abhängig von der Nachbarschaft erkannt werden. Jürgens und Eckhorn [1997] untersuchten die Fähigkeiten eines vollverbundenen Netzwerks aus *Marburger Modellneuronen*, die Korrelation bzw. den Korrelationskontrast von Eingangssignalen weiterzugeben. Dabei verglichen sie Varianten ohne, mit additiver und mit multiplikativer lateraler Kopplung. Es zeigte sich, daß Netzwerke ohne laterale Kopplung die Weitergabe sogar verschlechterten, während exzitatorische laterale Kopplung vorhandene Korrelationen im Input generell verstärkt. Interessant ist dabei, daß die multiplikative Kopplung in diesem Punkt bessere Ergebnisse brachte als ihr additives Gegenstück. Die Autoren führen dies darauf zurück, daß die nichtlineare Kennlinie der multiplikativen Kopplung eine 'strengere Korrelationsprüfung' implementiert als die lineare Kennlinie der additiven Kopplung. Die Unterschiede sind – ähnlich wie bei der Konturdetektion – vor allem deshalb nicht extrem, weil das Membranpotential zur Erzeugung des Ausgangssignals noch den Vergleich mit der Feuerschwelle durchlaufen muß. Diese stellt die hauptsächliche Nichtlinearität dar und sorgt dafür, daß sehr große und sehr kleine Membranpotentiale bei beiden Kopplungsvarianten einheitlich behandelt werden, nämlich als über- bzw. unterschwelliges Ausgangssignal. Die in Kap. 3.3.5 angestellten Überlegungen gelten hier sinngemäß.

7.5 Fazit

Aus den in dieser Arbeit vorgestellten Untersuchungen geht hervor, daß bereits eine rein datengetriebene Aufmerksamkeits- und Blicksteuerung bei geeigneter Vorverarbeitung in der Lage ist, in komplexen realen Szenen visuelle Objekte sinnvoll auszuwählen und ggf. in ihrer Bewegung zu verfolgen. Die enge Anlehnung an neurobiologische Erkenntnisse hat sich dabei als nützlich für die robuste Implementation einer solchen Steuerung erwiesen, zumal fast alle bisher bekannten technischen Systeme explizites Objektwissen voraussetzen. Diese Allgemeinheit wird mit einer geringeren Genauigkeit bei der Objektverfolgung erkauft; der modellbasierte und der modellfreie Ansatz sind als komplementär anzusehen und sollten für ein leistungsfähiges Gesamtsystem kombiniert werden.

Ein ähnliches Fazit läßt sich für die Szenensegmentierung ziehen: Während die biologisch orientierten Modelle prinzipiell in der Lage sind, mit komplexen realen Szenen umzugehen, kann die damit erzielte Segmentierung doch immer nur als ein 'erster Vorschlag'

angesehen werden, der durch eine genauere Analyse und Objektwissen zu ergänzen ist. Ob der vorgeschlagene Ansatz, den Grad der zeitlichen Korrelation in neuronaler Aktivität in einem geeignet strukturierten Netz direkt als 'Wahrscheinlichkeit der Zusammengehörigkeit' zu interpretieren, hier weiterführt, müssen zukünftige Untersuchungen zeigen.

Literaturverzeichnis

Amari, S. (1977). Dynamics of pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics*, 27:77–87. (document), 1.4.1, 3.5.2, 4, 4, 3.20, 3.5.8.4

Arndt, M. (1993). *Repräsentation räumlicher und zeitlicher Stetigkeit durch Synchronisation neuronaler Signale*. Dissertation, Philipps-Universität Marburg, FB Physik. 4.1.1

Arndt, M., Dicke, P., Erb, M., Eckhorn, R. und Reitböck, H. J. (1992). Two-layered physiology-oriented neuronal network models that combine dynamic feature linking via synchronization with a classical associative memory. In: *Neural Network Dynamics*, Taylor, J., Caianiello, E. und Cotterill, R., Hg., S. 140–154. Springer-Verlag. 4.2

Ballard, D. und Brown, C. (1982). *Computer Vision*. Prentice Hall. 1.4.3

Becker, W. und Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19:976–983. 1.4.1, 7.3.1

Burman, D. und Bruce, C. (1997). Suppression of task-related saccades by electrical stimulation in the primate's frontal eye field. *Journal of Neurophysiology*, 77:2252–2267. 7.3.1

Collewijn, H. (1972). An analog model of the rabbit's optokinetic system. *Brain Research*, 36:71–88. 7.3.2

Collewijn, H. (1981). *The oculomotor system of the rabbit and its plasticity*, Bd. 5 von *Studies of Brain Function*. Springer-Verlag, Berlin. 7.3.2

Collewijn, H. und van der Mark, F. (1972). Ocular stability in variable feedback condiditons in the rabbit. *Brain Research*, 36:47–57. 7.3.2

Collewijn, H. und Tamminga, E. (1986). Human fixation and pursuit in normal and open-loop conditions: effects of central and peripheral retinal targets. *Journal of Physiology*, 379:109–129. 7.3.2

Deubel, H. (1994). *Anatomy of a single refixation*. Habilitationsschrift, Ludwig-Maximilians-Universität, Müchen. 7.1

Dias, E. C. und Bruce, C. J. (1994). Physiological correlate of fixation disengagement in the primate's frontal eye field. *Journal of Neurophysiology*, 72(5):2532–2537. 7.3.1

Dodge, R. und Cline, T. (1901). The angle velocity of eye-movements. *Psychological Review*, 2:193–199. 1.2, 1.4.1

Dodge, R. und Fox, J. (1928). Optic Nystagmus. *Arch. Neurol. Psychiat.*, 20:812–823. 7.3.2

Dominey, P. und Arbib, M. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex*, 2:153–175. 7.3.1

Donders, F. (1847). Beitrag zur Lehre von den Bewegungen des menschlichen Auges. *Holländische Beiträge zu den anatomischen und physiologischen Wissenschften* , 1:104–145, 384–386. 1.4.1

Drüe, S., Hoischen, R. und Trapp, S. (1994). *Tolerante Objekterkennung durch das Neuronale Active-Vision-System NAVIS*, Bd. 5 von *Informatik Xpress*, S. 251–264. Technische Universität Wien. 8

Dubois, M. und Collewijn, H. (1979). Optokinetic Reactions in man elicited by localizes retinal motion stimuli. *Vision Research*, 19:1105–1115. 7.3.2

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. und Reitböck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130. 1.4.3

Eckhorn, R., Frien, A., Bauer, R., Woelbern, T. und Kehr, H. (1993). High frequency (60–90 Hz) oscillations in primary visual cortex of awake monkey. *NeuroReport*, 4:243–246. 1.4.3

Eckhorn, R., Reitböck, H. J., Arndt, M. und Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Computation*, 2:293–307. 1.1, 1.4.3, 2.6.1, 2.10, 3.3.3.1, 3.3.5, 6.3.2, 7.4.1

Ernst, U. (1993). . Diplomarbeit, Johann-Wolfgang-Goethe-Universität Frankfurt a.M. 4.2.2

van Essen, D. (1987). Visual Cortex, Extrastriate. In: *Encyclopedia of Neuroscience*, Adelman, G., Hg., Bd. II. Birkhäuser, 1. Aufl. 2.6

Faubert, J. und Grünau, von, M. (1995). The influence of two spatially distinct primers and attribute priming on motion induction. *Vision Research*, 35 (22):3119–3130. 2.5.4

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf & Härtel, Leipzig. 1.1

Fenske, J., Schott, U., Stöcker, M. und Reitboeck, H. (1995). Feature Contrast: I. A Neural Network for Pattern Segmentation via Motion Contrast Detection. In: *Proceedings of the 23rd Göttingen Neurobiology Conference 1995*, Elsner, N. und Menzel, R., Hg., Bd. 2, S. 889. Thieme Verlag, Stuttgart, New York. 3.4.1.2

Findlay, J. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*. 7.3.1

Findlay, J., Brogan, D. und Wenban-Smith, M. (1993). The spatial signal for saccadic eye movements emphasizes visual boundaries. *Perception and Psychophysics*, 53(6):633–641. 7.3.1

Findlay, J. und Gilchrist, I. (1997). Spatial scale and saccade programming. *Perception*, 26:1159–1167. 7.3.1

Fischer, B. und Ramsperger, E. (1984). Human express saccades: Extremely short reaction times of goal-directed eye movements. *Experimental Brain Research*, 57:191–195. 1.4.1

Frank, G., Hartmann, G., Jahnke, A. und Schäfer, M. (1996). An Accelerator for Neural Networks with Puls-Coded Model Neurons. *Submitted to the Journal of Artifical Neural Networks – Special issue on Pulse-Coupled Neural Networks.* 2.6.2, 2.11, 2.12, 3.2.4

French, A. und Stein, R. (1970). A Flexible Neural Analog Using Integrated Circuits. *IEEE BME*, 17(3):248–253. 1.1

Gabriel, A. und Eckhorn, R. (1999). Phase Continuity of Fast Oscillations may Support the Representation of Object Continuity in Striate Cortex of Awake Monkey - Correlation Analysis of Time- and Space-Resolved Single Responses. In: *Proceedings of the 27th Göttingen Neurobiology Conference*, Elsner, N. und Eysel, U., Hg., Bd. 2, S. 489. Georg Thieme Verlag. 1.4.3, 7.4

Gail, A., Brinksmeyer, H. und Eckhorn, R. (1999a). Different Possible Contributions of Striate Cortex Activity to Visual Object Representation in Awake Monkey. In: *Proceedings of the 27th Göttingen Neurobiology Conference*, Elsner, N. und Eysel, U., Hg., Bd. 2, S. 487. Georg Thieme Verlag. 1.4.3, 7.4

Gail, A., Brinksmeyer, H., Eckhorn, R. und Thomas, U. (1999b). Contributions of Spike-rate and Precise Correlations to Visual Object Representation in Striate Cortex of Awake Monkey. In: *submitted to SFN.* 7.4

Gerstner, W., Pith, R. und van Hemmen, J. L. (1991). Collective oscillations in the cortex: the importance of axonal transmission delays and postsynaptic response. 1.1, 7.4

Gerstner, W., Van Hemmen, J. L. und Cowan, J. D. (1996). What matters in neuronal locking? *Neural Computation*, 8:1653–1676. 4, 7.4

Gonzalez, R. und Woods, R. (1992). *Digital Image Processing*. Addison-Wesley. 1.4.3

Gottlieb, J. P., MacAvoy, M. G. und Bruce, C. J. (1994). Neural responses related to smooth-pursuit eye movements and their correspondence with electrically elicited smooth eye movements in the primate frontal eye field. *Journal of Neurophysiology*, 72(4):1634–1653. 7.3.1, 7.3.2, 7.2

Götzl, B. (1994). *Betriebsarten eines dynamischen Assoziativspeichers*. Diplomarbeit, Philipps-Universität Marburg, FB Physik. 4.2

Graefe, V. (1995). *Merkmalsextraktion, Objekterkennung, Echtzeit-Bildverarbeitungssysteme*, Bd. I, S. 121–192. infix-Verlag, Sankt Augustin. 1.4.2

Gray, C. und Singer, W. (1989). Stimuls-specific neuronal oscillations in orientation columns of cat viusal cortex. *Proc. Natl. Acad. Sci. USA*, 86:1698–1702. 1.4.3

Gray, C. M., Engel, A. K., König, P. und Singer, W. (1990). Stimulus-Dependent Neuronal Oscillations in Cat Visual Cortex: Receptive Field Properties and Feature Dependence. *European Journal of Neuroscience*, 2:607–619. 1.4.3

Groh, J. M., Born, R. G. und Newsome, W. T. (1997). How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *The Journal of Neuroscience*, 17(11):4312–4330. 7.3.1

v. Grünau, M., Racette, L. und Kwas, M. (1996a). Measuring the Attentional Speed-up in the Motion Induction Effect. *Vision Research*, 36:2433–2446. 2.5.4

v. Grünau, M., Racette, L. und Kwas, M. (1996b). Tow Contributions to Motion Induction: A preattentive Effect and Facilitation due to Attentional Capture. *Vision Research*, 36:2447–2457. 2.5.4

Guitton, D., Buchtel, H. und Douglas, R. (1985). Frontal lobe lesions in man cause difficulties in suppressing reflexive glances and in generating goal-directed saccades. *Experimental Brain Research*, 58:455–472. 7.3.1

Hanes, D., Patterson, W. und Schall, J. (1997). Role of the frontal eye field in countermanding saccades. *Journal of Neurophysiology*, 77:817–834. 7.3.1

Hartmann, G. (1982). Recursive features of circular receptive fields. *Biological Cybernetics*, 43:199–208. 3.2.3.3, 3.2.3.4, 3.4

Hartmann, G., Büker, U. und Drüe, S. (1999). *A hybrid neuro-AI-architecture*. Academic Press, San Diego. 7.2.1

von Helmholtz, H. (1866/1962). *Physiological Optics*. Dover. 1.4.1

Henriques, Y., Klier, E., Smith, M., Lowy, D. und Crawford, D. (1998). Gaze-centered remapping of remembered visual space in an open-loop pointing task. *Journal of Neuroscience*, 18(4):1583–1594. 7.3.1

Heywood, S. (1972). Voluntary control of smooth eye movements and their velocity. *Nature*, 238:408–410. 7.3.2

Hikosaka, O., Miyauchi, S. und Shimojo, S. (1993a). Focal visual attention produces illusory temporal order and motion sensation. *Vision Research*, 33 (9):1219–1240. 2.5.4, 2.9, 7.3.1

Hikosaka, O., Miyauchi, S. und Shimojo, S. (1993b). Visual attention revealed by an illusion of motion. *Neuroscience Research*, 18:11–18. 7.3.1

Hodgkin, A. L. und Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500–544. 1.1, 2.2

Hough, P. V. C. (1962). Method and means for recognizing complex patterns. US Patent 3.069.654. 1.4.3

Hubel und Wiesel (1962). Receptive Fields, binocular interaction and funtional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154. 3.3.1

Hughes, H. C., Nozawa, G. und Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, 8(3):197–230. 2.5.4

James, W. (1890). *The Principles of Psychology*. Macmillan, New York. 1.1

Jürgens, E. und Eckhorn, R. (1997). Parallel processing by a homogeneous group of coupled model neurons can enhance, reduce and generate signal correlations. *Biological Cybernetics*, 76:217–227. 7.4.1.1

Jürgens, E., Güttler und Eckhorn, R. (1999). Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-Potentials, but not in human EEG. *Experimental Brain Research*, 129:247–259. 7.4

Keating, E. und Gooley, S. (1988). Saccadic disorders caused by cooling the superior colliculus of the frontal eye fields or from combined lesions of both structures. *Brain Research*, 49:381–392. 7.3.1

Kistler, W. M., Seitz, R. und van Hemmen, J. L. (1998). Modeling collective excitations in cortical tissue. *Physica D*, 114:273–295. 7.4

Köhler, W. (1924). *Die physischen Gestalten in Ruhe und im stationären Zustand*. Verlag der Philosophischen Akademie, Erlangen. 1.1

Koller, D., Daniilidis, K. und Nagel, H. (1993). Model-based object-tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257–281. 1.4.2

Kopecz, K. (1995). Saccadic reaction times in gap/overlap paradigms: A model based on integration of intentional and visual information on neural, dynamic fields. *Vision Research*, 35:2911–2925. 1.4.1, 2.3.3, 3.5.2, 4, 4, 5, 3.20, 5, 3.5.4

Kopecz, K., Erlhagen, W. und Schöner, G. (1995). Dynamic representations provide the gradual specification of movement parameters. S. in press. 1.4.1

Kopecz, K. und Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural, dynamic fields. *Biological Cybernetics*, 73:submitted. (document), 1.4.1

Kowler, E. (1990). *The role of visual and cognitive processes in the control of eye movement*, Bd. 4 von *Reviews of Oculomotor Research*, Kap. 1. Elsevier, Amsterdam, New York, Oxford. 7.3.1, 7.3.2

Kowler, E. und McKee, S. (1987). Sensitivity of smooth eye movements to small differences in target velocity. *Vision Research*, 27:993–1015. 7.3.2

Kowler, E., van der Steen, J., Tamminga, E. und Colleweijn, H. (1984). Voluntary selection of the target for smooth eye movement in the presence of superimposed, full-field stationary and moving stimuli. *Vision Research*, 24:1789–1798. 7.3.2

Kowler, E. und Steinman, R. (1979a). The effect of expectations on slow oculomotor control. I. Periodic target steps. *Vision Research*, 19:619–632. 7.3.2

Kowler, E. und Steinman, R. (1979b). The effect of expectations on slow oculomotor control. II. Single target displacements. *Vision Research*, 19:633–646. 7.3.2

Kowler, E. und Steinman, R. (1981). The effect of expectations on slow oculomotor control. III. Guessing unpredictable target displacements. *Vision Research*, 21:191–203. 7.3.2

Leuck, H. und Nagel, H.-H. (1999). Automatic Differentiation Facilitates OF-Integration into Steering-Angle-Based Road Vehicle Tracking. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99), Fort Collins/CO, 23-25 June 1999*, Bd. 2, S. 360–365. IEEE Computer Society Press, Los Alamitos/CA 1999. 1.1

Lindemann, J., Kopecz, K. und Eckhorn, R. (1998). Model-Based tracking of visual objects with learning neural networks. In: *Proceedings of the 26th Göttingen Neurobiology Conference*, Elsner, N. und Wehner, R., Hg., Bd. 2, S. 768. Thieme Verlag. 7.2.1

Lüschow, A. und Nothdurft, H. (1993). Pop-Out of Orientation but no Pop-Out of Motion at Isoluminance. *Vision Res.*, 33(1):91–104. 2

Lüke, H. (1975). *Signalübertragung*. Springer-Verlag. 3.2.3.2

MACKEBEN, M. und NAKAYAMA, K. (1993). Express attentional shifts. *Vision Research*, 33(1):85–90. 7.3.1

VON DER MALSBURG, C. und SCHNEIDER, W. (1986). A Neural Cocktail-Party Processor. *Biological Cybernetics*, 54:29–40. 1.4.3

MAYS, L. und SPARKS, D. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, 43:207–232. 7.3.1

MCCULLOCH und PITTS (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133. 2.6.3, 2.13

MCGOWAN, J., KOWLER, E., SHARMA, A. und CHUBB, C. (1998). Saccadic localization of ramdom dot targets. *Vision Research*, 38:895–910. 3.5.5.1

METZGER, W. (1936). *Gesetze des Sehens*. Kramer, Frankfurt. 1.1

MOHRAZ, K., SCHOTT, U. und PAULY, M. (1997). Parallel Simulation of Pulse Coded Neural Networks. In: *Proceedings of the 15th IMACS World Congress*, Bd. 6, S. 523–528. 1.2

MÖLLER, M. (1995). *Leistungsvergleich verschiedener Neuronenmodelle bei grundlegenden Aufgaben der Informationsverarbeitung*. Diplomarbeit, Philipps-Universität Marburg, FB Physik. 3.2.4, A.1

MUNOZ, D. P. und WURTZ, R. H. (1992). Role of the Rostral Superior Colliculus in Active Visual Fixation and Execution of Express Saccades. *Journal of Neurophysiology*, 67:1000–1002. 1.4.1, 7.3.1

MUNOZ, D. P. und WURTZ, R. H. (1993a). Fixation Cells in Monkey Superior Colliculus I. Characteristics of Cell Discharge. *Journal of Neurophysiology*, 70:559–575. 1.4.1, 3.5.7, 7.3.1

MUNOZ, D. P. und WURTZ, R. H. (1993b). Fixation Cells in Monkey Superior Colliculus II. Reversible Activation and Deactivation. *Journal of Neurophysiology*, 70:576–589. 1.4.1, 7.3.1

MUNOZ, D. P. und WURTZ, R. H. (1995a). Saccade-related activity in monkey superior colliculus. I.Characteristics of burst and build-up cells. *Journal of Neurophysiology*, 73:2313–2333. 1.4.1, 7.3.1

MUNOZ, D. P. und WURTZ, R. H. (1995b). Saccade-related activity in monkey superior colliculus. II. Spread of activity during saccades. *Journal of Neurophysiology*, 73:2334–2348. 1.4.1, 7.3.1

NAGEL, H. (1985). Analyse und Interpretation von Bildfolgen. *Vision Res.*, 33(1):178–200. 1.4.2

Newsome, W. T., Wurtz, R. H., Dürsteler, M. R. und Mikami, A. (1985). Deficits in Visual Motion Processing Following Ibotenic Acid Lesions of the Middle Temporal Visual Area of the Macaque Monkey. *The Journal of Neuroscience*, 5:825–840. 7.3.1

Nischwitz, A. und Glünder, H. (1995). Local lateral inhibition: a key to spike synchronization? *Biological Cybernetics*, 73:389–400. 4.2.2, 7.4

Opara, R. und Wörgötter, F. (1996). Using Visual Latencies to Improve Image Segmentation. *Neural Computation*, 8:1493–1520. 4, 4.4.1

Pauly, M., Kopecz, K. und Eckhorn, R. (1997). A dynamic gaze control model network of pulse-coding neurons: concurrent treatment of saccades and smooth pursuit. In: *Proceedings of the 25th Göttingen Neurobiology Conference*, Elsner, N. und Wässle, R., Hg., Bd. 2, S. 1022. Thieme Verlag. 1.2

Pauly, M., Kopecz, K. und Eckhorn, R. (1998). Lateral Coupling Preserves Object Contours by Reducing 'Natural' Types of Noise in Models of Orientation Detector Maps. In: *Proceedings of the 26th Göttingen Neurobiology Conference*, Elsner, N. und Wehner, R., Hg., Bd. 2, S. 769. Thieme Verlag. 1.2

Pauly, M., Kopecz, K. und Eckhorn, R. (1999). Gaze control with neural networks: A unified approach for saccades and smooth pursuit. In: *Engineering Applications of Bio-Inspired Artificial Neural Networks. Proceedings of the International Work-Conference on Artificial and Natural Neural Networks, IWANN'99*, S. 1022. Springer Verlag. 1.2, 8

Pierrot-Deseilligny, C., Rivaud, S., Gaymard, B. und Agid, Y. (1991). Cortical control of reflexive visually guided saccades. *Brain*, 114:1473–1485. 7.3.1

Pierrot-Deseilligny, C., Rivaud, S., Gaymard, B., Müri, R. und Vermersch, A.-I. (1995). Cortical control of saccades. *Ann. Neurol.*, 37:557–567. 7.3.1

de Pucket, J. und Steinman, R. (1969). Tracking eye movements with and without saccadic correction. *Vision Research*, 9:695–703. 7.3.2

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *Journal of Neurophysiology*, 159:326–338. 7.3.2

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems (bei der optischen Wahrnehmung eines Insekts). *Z. Naturforschung*, 12:448–475. 3.4.1.1, 3.17

Reitböck, H. J. und Altmann, J. (1984). A model for size- and rotation-invariant pattern processing in the visual system. *Biological Cybernetics*, 51:113–121. 3.2.3.5, 7.2.1

Rieke, F., Warland, D., de Ruyter van Stevenick, R. und Bialek, W. (1999). *Spikes – Exploring the neural code*. MIT Press. 7.4.1

Ritz, R., Gerstner, W., Fuentes, U. und van Hemmen, J. L. (1994). A biologically motivated and analytically soluble model of collective oscillations in the cortex II. Application to binding and pattern segmentation. *Biological Cybernetics*, 71:349–358. 7.4

Robinson, D. (1972). Eye Movements evoked by collicular Stimulation in the alert Monkey. *Vision Research*, 12:1795–1808. 1.4.1

Robinson, D. und Fuchs, A. (1969). Eye movements evoked by stimulation of the frontal eye fields. *J. Neurophysiol.*, 32:637–648. 1.4.1

Roth, G. und Prinz, W. H. (1996). *Kopf-Arbeit*. Spektrum Akad. Verlag, Heidelberg. 2.1, 2.3, 2.5

Saam, M. (1999). Pers. Kommunikation. 7.4

Saam, M., Eckhorn, R. und Schanze, T. (1999). Spatial Range of Synchronization in Visual Cortex is Determined by Lateral Conduction Velocity and Determines RF-Size at Next Processing Level. In: *submitted to SFN*. 7.4

Saslow (1967). Latency for saccadic eye movement. *Journal of the American Optical Society*, 57:1030–1033. 1.4.1

Schanze, T. und Eckhorn, R. (1997). Phase correlation among rythms present at different frequencies: spectral methods, application to microelectrode recordings from visual cortex and functional implications. *International Journal of Psychophysiology*, 26:171–189. 2.4

Schierwagen, A. (1996). *Gaze control in active vision: neural field model of dynamic motor error coding*, Bd. 4, S. 46–51. infix-Verlag, Sankt Augustin. 3.5.7.1

Schiller, P. und Chou, I.-H. (1998). The effects of frontal eye field and dorsomedial frontal cortex lesions on visually guided eye movements. *Nature Neuroscience*, 1:248–253. 7.3.1

Schmidt, R. und Thews, G. H. (1996). *Physiologie des Menschen*. Springer Verlag, Heidelberg Berlin New York Tokyo, 26. Aufl. 2.4, 2.7

Schneider, J., Eckhorn, R. und Reitböck, H. J. (1983). Evaluation of neuronal coupling dynamics. *Biological Cybernetics*, 46:129–134. 1.4.3

Schott, U. (1999). *Entwicklung eines impulskodierenden neuronalen Netzes für die Segmentierung bewegter Szenen*. Dissertation, Philipps-Universität Marburg, FB Physik. 1, 3.4, 3.18, 3.4.1.2, 3.19, A.1, A.2

Spengler, C. (1996). *Relationale Kodierung von Objektmerkmalen für die visuelle Mustererkennung in biologienahen neuronalen Netzen*. Diplomarbeit, Philipps-Universität Marburg, FB Physik. 3.2.3.3, 3.1, 7.4

Stöcker, M. (1993). *Höhere Mechanismen der Informationsverarbeitung*. Dissertation, Philipps-Universität Marburg, FB Physik. 7.4

Stöcker, M. (1994). Persönliche Kommunikation. 3.2.3.2

Ter Braak, J. (1936). Untersuchungen über den optokinetischen Nystagmus. *Arch. Neerl. Physiol.*, 60:131–135. 7.3.2

Ter Braak, J. (1957). 'Ambivalent' optokinetic stimulation. *Fol. Psychiat. Neurol. Neurochir. Neerl.*, 60:131–135. 7.3.2

Ter Braak, J. und Buis, C. (1970). Optokinetic Nystagmus and attention. *International Journal of Neurophysiology*, 8:34–42. 7.3.2

Wallach (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychol. Forschung*, 20:325–380. 3.4.1.1

Weitzel, L. (1998a). *Bildsegmentierung durch Texturanalyse mit neuronalen Netzen*. Diplomarbeit, Philipps-Universität Marburg, FB Physik. 7.4

Weitzel, L. (1998b). *Extraktion und Separation von Konturen und Formen mit dynamischen Neuronalen Netzen in realen Szenen*. Dissertation, Philipps-Universität Marburg, FB Physik. 1, 3.3, 3.3.2, 3.10, 3.11, 5.2, 7.4

Weitzel, L., Kopecz, K., Spengler, C. und Eckhorn, R. (1997). Contour Based Figure-Ground Segregation fo Real World Images Modelled by a Linking Architecture of Pulse-Coding Neurons. In: *Proceedings of the 25th Göttingen Neurobiology Conference*, Bd. 2, S. 1027. 7.4

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 61. 1.1, 1.4.2

Westheimer, G. (1954a). Eye movements responses to a horizontally moving stimulus. *Archives of Ophtalmology*, 52:932–941. 1.4.1

Westheimer, G. (1954b). The mechanism of saccadic eye movements. *Archives of Ophtalmology*, 52:710–714. 1.4.1

Young, R. (1986). The gaussian derivative model for machine vision: visual cortex simulation. *Journal of the optical society of America*, GMR-5323. 3.2.3.4

Zell, A. (1994). *Simulation neuronaler Netze*. Addison-Wesley. 3.2.4

Anhang A

Details zur Simulationstechnik

A.1 Vorverarbeitung

Wie in Kap. 2.6 bereits beschrieben, wurden alle Netzwerke mit Marburger Modellneuronen in der kompakten Beschreibungssprache MNET formuliert. Eine vollständige Beschreibung von MNET mit Beispielen findet sich in [Möller, 1995]; dem Leser mit Kenntnissen in C sollte sich die Syntax jedoch leicht erschließen.

Für die compiler-technisch interessierten Leser sei noch angemerkt, daß die in den folgenden Beispieldateien erscheinenden C-Ausdrücke tatsächlich vom MNET-Compiler mit weiteren Standard-Bausteinen zu einem vollständigen C-Programm zusammengesetzt werden, das anschließend normal kompiliert und ausgeführt wird. Bei der Ausführung wird eine Binärbeschreibung der Netztopologie (d.h. eine Verbindungstabelle) erzeugt, die gleichermaßen als Input für den Hardware-Accelerator wie für seine Software-Simulation dient. Dieses zweistufige Verfahren erlaubt zum einen eine Trennung von Netzbeschreibung und Simulationsalgorithmus, zum anderen gibt es durch die unmittelbare Verwendung von C-Ausdrücken zur Beschreibung von Verschaltungsmustern und Neuronenschichten vollständige Freiheit in der Formulierung der Netzwerktopologie.

Das Einlesen der Eingangssequenz und die Simulation der in Kap. 3.2 beschriebenen Vorverarbeitung (X-Zellen) erfolgt in einem getrennten Programm und erzeugt für jeden Zeitschritt eine Spikeliste als Ausgabe (Nummern der spikenden Neuronen als ASCII-Datenstrom, vgl. auch Abb. 3.1). Zusätzlich verfügt es über eine Schnittstelle, die es ermöglicht, Bewegungsanweisungen von der Blicksteuerung entgegenzunehmen. Das Verhalten des X-Zellen-Programms wird von der Datei `para.dat` gesteuert, die später per `#include`-Befehl in die eigentliche Netzbeschreibung eingebunden wird:

```
//InputnameRumpf=/usr/ax1315/elauge/bildmaterial/bmw/dt/dt
//BMP_OutputnameRumpf=/usr/ax1303/pauly/elauge
//Output-Verzeichnis=/usr/ax1303/pauly/elauge/aufmerk/dt_sprung12
ASCII=0;
BREITE=220;
HOEHE=191;
Karlsruhe=1;
Digits=2;
```

```
Startbild=10;
Frames=50;
Gap=32;
Aufloesungen=2;
AbtastpunkteX=32;
AbtastpunkteY=22;
FixpunktX=150;
FixpunktY=86;

//Aufloesung0
Abtastweite0=4;
Size0=10;
DS_Inc_ON0=58;
DS_Inc_OFF0=58;
DS_Tau_ON0=15;
DS_Tau_OFF0=15;
DS_Offset_ON0=10;
DS_Offset_OFF0=10;

//Aufloesung1
Abtastweite1=2;
Size1=4;
DS_Inc_ON1=58;
DS_Inc_OFF1=58;
DS_Tau_ON1=15;
DS_Tau_OFF1=15;
DS_Offset_ON1=10;
DS_Offset_OFF1=10;
```

Die mit `//` beginnenden Zeilen sind Pseudo-Kommentare, die vom X-Zellen-Programm ausgewertet, vom MNET-Compiler aber überlesen werden. Die Angaben in den Abschnitten `Aufloesung`n bestimmen die Parameter der X-Zellen für die verwendeten Auflösungsstufen (für eine detaillierte Beschreibung s. [SCHOTT, 1999]). Die Angaben `Abtastweite`n`=`m und `Size`n`=`m beziehen sich dabei auf die in Kap. 3.2 angegebenen Größen d und M für Abtastweite und Maskengröße. `FixpunktX` und `FixpunktY` geben die Position der anfänglichen Blickrichtung, d.h. die Mitte des Gesichtsfeldes, in Pixel-Koordinaten an. `BREITE` und `HOEHE` replizieren lediglich die Ausmaße der verwendeten Eingangsbilder in Pixeln. `Gap` entspricht der Größe Δt_{Frame} aus Kap. 3.2.

A.2 Netzbeschreibung in MNET

Da das Lesen fremder Quelltexte nach meiner Erfahrung eine eher mühsame Angelegenheit darstellt (und deshalb meistens gar nicht erst stattfindet), gebe ich im folgenden beispielhaft die kommentierte Netzbeschreibung aus der ersten Beispielszene wieder (Simulationsbeispiel 1a, vgl. Kap. 5.1.1). Dieses Netz implementiert die in Kap. 3.5 be-

schriebene Verschaltungsstruktur für die Aufmerksamkeitsschicht, zusammen mit dem pseudo-hexagonalen Gitter für die Neuronenanordnung.

In den weiteren Simulationsbeispielen zur Verfolgung (1a, 2, 3a und 3b) wurde an Stelle des X-Zellen-Outputs die Ausgabe der Bewegungskontrast-Detektion von SCHOTT [1999] als Input für die Aufmerksamkeitsschicht verwendet. Da das Transientensystem bereits alle im Hardware-Accelerator verfügbaren Zeitkonstanten belegt, wurde für die Aufschaltung der Bewegungskontrast-Detektoren auf die Aufmerksamkeitsschicht (und damit auch für die interne Wechselwirkung der Aufmerksamkeitsschicht) mit 15 bin die längste dort verwendete Zeitkonstante übernommen.

Beispielszene 2 (Fußgängerin) ist erheblich besser aufgelöst als die anderen Szenen, was einerseits dazu führt, daß Objekte (in Pixel gemessen) größer erscheinen und andererseits die Simulationszeiten durch die große Zahl von Neuronen und Verbindungen sehr in die Höhe getrieben werden. Um diesen Problemen gerecht zu werden, wurde hier nur die Auflösung 2 mit Abtastweite $d=4$ verwendet und die Breite σ der internen Verschaltung der Aufmerksamkeitsschicht auf 20 Pixel vergrößert.

Die Zeitkonstanten der Synapsen werden aus Gründen der Anpassung an den Hardware-Accelerator in einer separaten Tabelle verwaltet; die `lookup`-Einträge in der Netzbeschreibung verweisen auf diese Tabelle. Um das Verständnis der Netzbeschreibung einfacher zu gestalten, sind die verwendeten Werte direkt bei der Definition der Synapsen mit angegeben. Die Schwellenzeitkonstante betrug für alle Neurone $\tau_\Theta = 5\,\text{bin}$, der Schwellenoffset $\Theta_0 = 8$.

```
/****************** Netzwerkbeschreibung Verfolgungsnetz ******************
 ***              Einfache Version: X-Zellen sind direkt                ***
 ***              auf die Aufmerksamkeitsschicht aufgeschaltet          ***
 ***************************************************************************/

network Verfolgungsnetz;

defaults
learnfunc=Std_Backpropagation;
updatefunc=Synchronous_Order;

//zunächst einige Hilfs-Makros
#define MaxAbtast Abtastweite0  //groebste Aufloesung muss Nummer 0 haben!!

//Position der linken oberen Ecke für die Neuronenschichten berechnen
#define Links0vor  (FixpunktX-((Abtastweite0 / 2)*(AbtastpunkteX)))
#define Links0     (Links0vor + MaxAbtast - (int)Links0vor%(int)MaxAbtast)
#define Oben0vor   (FixpunktY-((Abtastweite0 / 2)*(AbtastpunkteY)))
#define Oben0      (Oben0vor + MaxAbtast - (int)Oben0vor%(int)MaxAbtast)

#define Links1vor  (FixpunktX-((Abtastweite1 / 2)*(AbtastpunkteX)))
#define Links1     (Links1vor + MaxAbtast - (int)Links1vor%(int)MaxAbtast)
#define Oben1vor   (FixpunktY-((Abtastweite1 / 2)*(AbtastpunkteY)))
#define Oben1      (Oben1vor + MaxAbtast - (int)Oben1vor%(int)MaxAbtast)
```

```
//MNET-Abschnitt für Konstanten
const

//Einbinden der Beschreibungsdatei der Vorverarbeitung
//Damit kennnen wir Größe und Anordnung der Neuronenschichten, s.o.
#include <para.dat>

// Feeding-Gewicht, mit dem die X-Zellen
// auf die Aufmerksamkeitsschicht aufgeschaltet sind (Synapse Feeding1)
// Da dies der eigentliche Input ist, wird hier eine lange Zeitkonstante
// verwendet: 30 bin
feed_aufm_0 = 0.7;
feed_aufm_1 = 0.2;

radius_aufm_0 = 8;
radius_aufm_1 = 8;

// Parameter für die interne Verschaltung der Aufmerksamkeitsschicht
// (Gaußförmige Vollverbindung mit Breite Sigma und negativem Offset NegOffset)
GaussFaktor = 0.7;
Sigma = 5.66;
NegOffset = 0.5;

// Aufgrund eines Fehlers im MNET-Compiler muß die Division
// als Multiplikation mit dem Kehrwert durchgeführt werden, deshalb
// hier noch der Kehrwert:
KWSigma = 0.565;
NegOffsetFixpunkt = 0.2;

// Nach einer Sakkade treten oft starke Einschalteffekte an den Rändern auf,
// deshalb sorgen wir dafür, daß die Gegend um den Fixpunkt kurzzeitig bevorzugt wird
// Zeitkonstante (Feeding2): 8 bin
// Parameter für zusätzliche Verschaltung zum Fixpunkt
GaussFaktorFixpunkt = 0.42;
SigmaFixpunkt = 14.1;
KWSigmaFixpunkt = 0.0707;

// MNET-Typdefinitionen für:
// - Synapsen (Sites)
// - Neuronentypen (Units)
// - Neuronenschichten (Maps)
typedef

// zunächst die Synapsen des Marburger Modellneurons
site with sitefunc Site_WeightedSum : Linking;
site with sitefunc Site_WeightedSum : Feeding1;
```

```
site with sitefunc Site_WeightedSum : Feeding2;
site with sitefunc Site_WeightedSum : InhSite;

// Neuronentypen für die unterschiedlichen Auflösungsstufen,
// hier alle identisch
// Die lookup-Einträge verweisen auf die Tabelle der Zeitkonstanten
// Linking- und inhibitorische Synapse werden für die
// Aufmerksamkeitsschicht nicht benötigt.
// Als Zeitkonstanten wurden verwendet:
// Feeding 1: 30 bin
// Feeding 2:  8 bin
// Schwelle :  5 bin

unit with name "StandardNeuron"
                sites Feeding1, Feeding2, Linking, InhSite
                lookup 1,1,1,1,1
                : Standard0;

unit with name "StandardNeuron"
                sites Feeding1, Feeding2, Linking, InhSite
                lookup 1,1,1,1,1
                : Standard1;

unit with name "StandardNeuron"
                sites Feeding1, Feeding2, Linking, InhSite
                lookup 1,1,1,1,1
                : Standard2;

// Layout-Definitionen für die verwendeten Auflösungsstufen
// (als Boolsche Ausdrücke in C-Syntax)
// Zunächst die rechteckigen Ausschnitte, unter Verwendung der Angaben aus para.dat
mapdef ausschnitt0 = (x >= (Links0) && x < (Links0+AbtastpunkteX*Abtastweite0)
                     && y >= (Oben0) && y < (Oben0+AbtastpunkteY*Abtastweite0));
mapdef ausschnitt1 = (x >= (Links1) && x < (Links1+AbtastpunkteX*Abtastweite1)

// Nun die hexagonalen Raster für die verwendeten Auflösungsstufen
mapdef hex0 = ( ((int)y % (int)(2*Abtastweite0) == 0
                     && (int)x % (int)Abtastweite0 == 0 )
                  || ((int)y % (int)(2*Abtastweite0) == Abtastweite0
                     && (int)x % (int)Abtastweite0 == Abtastweite0 / 2) );
mapdef hex1 = ( ((int)y % (int)(2*Abtastweite1) == 0
                     && (int)x % (int)Abtastweite1 ==0 )
                  || ((int)y % (int)(2*Abtastweite1) == Abtastweite1
                     && (int)x % (int)Abtastweite1 == Abtastweite1 / 2) );

// Logisches verUNDen liefert das eigentliche Schichten-Layout:
// hexagonales Raster in der gewünschten Ausdehnung
```

```
mapdef raster0 = ( ausschnitt0 && hex0 );
mapdef raster1 = ( ausschnitt1 && hex1 );
mapdef raster2 = ( ausschnitt2 && hex2 );

// Damit können wir die eigentlichen Neuronenschichten anlegen.
// Zunächst die standardisierten Schichten für die Vorverarbeitung,
// jeweils ein Satz X-ON und X-OFF-Neurone für jede Auflösung:
cluster[BREITE, HOEHE] of Standard0 layout raster0
         iotype input
         bias 16
         : XON0, XOFF0;

cluster[BREITE, HOEHE] of Standard1 layout raster1
        iotype input
        bias16
        : XON1, XOFF1;

cluster[BREITE, HOEHE] of Standard2 layout raster2
        iotype input
        bias16
        : XON2, XOFF2;

// Nun die eigentliche Aufmerksamkeitsschicht
cluster[BREITE, HOEHE] of Standard0 layout raster1
        iotype hidden
        bias 16
        :AUFM;

// Als nächstes folgt die Beschreibung der Verschaltungsmuster mit dem connection-Befehl.
// Die Bezeichner x,y,tx und ty sind reserviert und stehen für die Koordinaten des
// Quellneurons (x,y) und des Zielneurons (tx,ty) einer Verbindung. Das Ergebnis des
// Ausdrucks liefert für jede Verbindung zweier Neurone das synaptische Gewicht,
// wobei der reservierte Wert ‘none’ ‘keine Verbindung’ bedeutet.

// Quadratischer Abstand zweier Neurone
connection DistSq = ((x-tx)*(x-tx)+(y-ty)*(y-ty));

// Abstand zweier Neurone
connection Dist = (sqrt(DistSq));
connection DistFixpunktSq
       = ((tx-FixpunktX)*(tx-FixpunktX)+(ty-FixpunktY)*(ty-FixpunktY));

// Quadratische Breite der Gauss-Funktionen
connection KWSigmaSq = (KWSigma*KWSigma);
connection KWSigmaFixpunktSq = (KWSigmaFixpunkt*KWSigmaFixpunkt);

// Ein räumlich homogenes Gaußprofil mit Breite Sigma und Maximum GaussFaktor
```

```
connection Gauss = GaussFaktor*(exp(-0.5*DistSq*KWSigmaSq));

// Dasselbe, aber zentriert auf den Fixpunkt
connection GaussAufFixpunkt
       = GaussFaktorFixpunkt*(exp(-0.5*DistFixpunktSq*KWSigmaFixpunktSq));

// Jede X-Zelle erregt einen kreisförmigen Bereich der Aufmerksamkeitsschicht
// um die eigene Position herum ( (Bedingung ? dann_zweig : sonst_zweig)
// ist in MNET, wie in C, eine Kurzform für IF ... THEN ... ELSE...)
connection x02aufm = ( (Dist <= radius_aufm_0) ? (feed_aufm_0) : none);
connection x12aufm = ( (Dist <= radius_aufm_1) ? (feed_aufm_1) : none);

connection aufm2aufm = (Gauss-NegOffset);
connection aufm2Fixpunkt = (GaussAufFixpunkt-NegOffsetFixpunkt);

// Im Abschnitt 'topology' wird die eigentliche Verschaltung veranlaßt.
// Dabei wird potentiell jedes Neuron der Quellschicht mit jedem Neuron der
// Zielschicht verbunden, entsprechend dem mit 'weight' aufgerufenen Muster:
topology

// Verbinde X-ON-Zellen der Auflösung 0 mit der Aufmerksamkeitsschicht
// Benutze dazu das Muster x02aufm
// Da dies der eigentliche Input ist, wird hier eine lange Zeitkonstante
// verwendet: 30 bin
connect XON0 with AUFM
        through Feeding1
        weight  x02aufm;

// Nun die x-OFF-Zellen
connect XOFF0 with AUFM
        through Feeding1
        weight  x02aufm;

// ... und das gleiche für Auflösung 1
connect XON1 with AUFM
        through Feeding1
        weight  x12aufm;

connect XOFF1 with AUFM
        through Feeding1
        weight  x12aufm;

// Verbinde die Aufmerksamkeitsschicht gemäß Gl. 3.14
// Auch dieser Input verwendet die lange Zeitkonstante
// von Feeding1: 30 bin
connect AUFM with AUFM
        through Feeding1
```

```
        weight  aufm2aufm;

// Die transiente Anregung des Fixpunkts erfolgt mit der Zeitkonstante
// von Feeding2: 8 bin
connect AUFM with AUFM
        through Feeding2
        weight  aufm2Fixpunkt;

// Zum Schluß noch die Numerierung der Schichten zur
// besseren Identifikation bei der Visualisierung
locate XON0 at 0,0,1;
locate XOFF0  at 0,0,2;
locate XON1 at 0,0,3;
locate XOFF1  at 0,0,4;
// locate XON2 at 0,0,5;
// locate XOFF2  at 0,0,6;
locate AUFM at 0,0,7;

end
```

Für den Verfolgungsparameter $\hat{\alpha}$ wurde in allen Simulationen ein Wert von 0.2 gewählt, eine Sakkade wurde für Abweichungen größer als $K_{Sacc} = 7$ Pixel ausgelöst.

A.3 Auswertung des Aufmerksamkeitssignals und Kommunikation zwischen den Modulen

Die Vorverarbeitung und die eigentliche Netzsimulation werden über eine UNIX-*Shell-Pipe* verbunden, d.h. die einzelnen Module schreiben nach *stdout*, alle Module außer der Vorverarbeitung lesen von *stdin*. Diese Technik ermöglicht mit einfachen Mitteln einen modularen Aufbau des Systems, verbunden mit der Möglichkeit, weitere Module zur Auswertung oder auch Veränderung der Spikefolgen einzufügen. Davon wurde z.B. bei der Einführung der aufmerksamkeitsabhängigen Latenzen Gebrauch gemacht (s. Kap. 4.4).

Zur Ermittlung der Blickzielposition wird das retinotope Signal der Aufmerksamkeitsschicht gemäß Gl. 3.17 ausgewertet, d.h. die Spikes werden jeweils über $\Delta t_{Move} = 8$bin integriert und anschließend der Schwerpunkt berechnet. Die so gewonnene Targetposition wird mit der momentanen Blickrichtung verglichen und die fällige Bewegung in horizontaler und vertikaler Richtung der Vorverarbeitung über die o.g. Schnittstelle mitgeteilt. Dabei handelt es sich um eine direkt programmierte *Socket*-Verbindung, auf der in jedem Zeitschritt ein Befehl übergeben wird (keine Bewegung wird durch $dx = 0$, $dy = 0$ angezeigt, vgl. auch Abb. 3.20a und Gl. (3.19)). Damit ist der synchronisierte Betrieb in einer geschlossenen Schleife möglich.

A.4 Simulationen zur Segmentierung

Bei den Simulationen zur Objekt-Hintergrund-Segmentierung (Kap. 4 und Kap. 5.2, Beispiele 4a,b) wurden lange Zeitkonstanten benötigt, was den gleichzeitigen Betrieb mit den Transientensystem ausschloß. Deshalb wurden für die Beispiele 4a und 4b aus Kap. 5.2 die Trajektorie der Verfolgung aus Beispiel 1 in die Vorverarbeitung eingespeist und so dem Segmentierungsnetz eine Input-Sequenz mit simulierter Kamerabewegung präsentiert. Die Parameter des Segmentierungsnetzes (vgl. Kap. 4.2.1) sind in Tab. A.1 angegeben.

Tabelle A.1: Parameter der Liniendetektoren im Segmentierungsnetz

V_F	V_L	V_I	V_Θ	Θ_0	τ_F	τ_L	τ_I	τ_Θ
1.4	1.5	2.5	24	5	10 bin	2 bin	20 bin	40 bin

Danksagung

Diese Dissertation hat eigentlich zwei Doktorväter: Prof. Dr. Dr. H. J. Reitböck übergab bei seiner Emeritierung im Juni 1998 die Betreuung an Prof. Dr. R. Eckhorn, der die Arbeit schon vorher mit großem Engagement und Interesse begleitet hatte. Dafür und für die Möglichkeit, diese Arbeit in der AG Neurophysik anzufertigen, möchte ich mich bei beiden Professoren herzlich bedanken. Prof. Dr. F. Rösler danke ich dafür, daß er sich auf unkomplizierte und flexible Weise bereiterkärte, das Zweitgutachten zu übernehmen. Des weiteren danke ich Herrn Dr. J. Radloff für die ausdrückliche Unterstützung meiner wissenschaftlichen Arbeit, während ich bereits Mitarbeiter des Hochschulrechenzentrums war, ebenso allen anderen Kollegen aus dem HRZ.

Die aufgeschlossene Arbeitsatmosphäre und der gute Zusammenhalt ist ein Verdienst der gesamten AG Neurophysik. Besonders danke ich aber Mirko Saam und Andreas Gabriel für viele engagierte Diskussionen rund um neuronale und andere Computer und ihre ständige tatkräftige Hilfsbereitschaft bei Problemen damit. Ulrich Schott, Lothar Weitzel, Karim Mohraz, Michael Erb und Klaus Kopecz gilt mein Dank für die mehr als kollegiale Zusammenarbeit, besonders im 'Elektronischen Auge'. Die beiden letztgenannten wirkten mit wissenschaftlicher Erfahrung und persönlichem Einsatz an entscheidenden Punkten meiner Arbeit klärend und ermutigend. Sigrid Thomas nahm die Mühe des Korrekturlesens auf sich und wußte schnellen und kompetenten Rat in allen organisatorischen Fragen.

Martin Jung danke ich für herausragende Leistungen in der Kochkunst und geduldiges Ertragen von gelegentlichem Chaos im Haus.

Meinen Eltern danke ich für die Ermöglichung eines interessanten und erfüllenden Physikstudiums sowie für ihre großzügige Unterstützung, die sie mir in jeder Hinsicht gewährten.

Meine Eltern ermöglichten mir nicht nur ein interessantes und erfüllendes Physikstudium, sondern ließen mir auch sonst in jeder Hinsicht große Unterstützung zukommen.

Karin Troidl schließlich ist die einzige mir bekannte Person, die scharfen Verstand und köstlich-schrägen Humor so gekonnt in sich vereinigt, daß ohne sie mein Leben anders verlaufen wäre. Ich schulde ihr ebenfalls Dank für viele anregende Diskussionen rund um Augenbewegungen und für große Stapel relevanter Literatur.

Eidesstattliche Versicherung

Ich versichere, daß ich meine Dissertation *Ein aufmerksamkeitsgestütztes, biologienahes Objekt-Erkennungs- und Verfolgungssystem mit impulscodierenden Neuronen* selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 2. März 2000